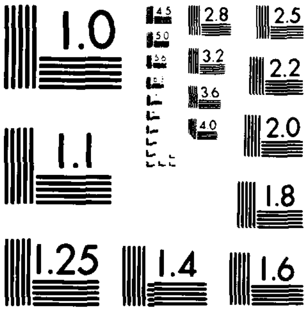
AD-A083 824 WISCONSIN UNIV-MADISON MATHEMATICS RESEARCH CENTER F/G 12/1
HIGH ORDER DIFFERENCE METHODS FOR QUASILINEAR ELLIPTIC BOUNDARY--ETC(U)
FEB 80 K BOEHMER DAAG29-75-C-0024
UNCLASSIFIED MRC-TSR-2042 NL

AD
A083824

END
DATE
FILMED
6-80
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

MRC Technical Summary Report #2042

HIGH ORDER DIFFERENCE METHODS FOR QUASILINEAR ELLIPTIC BOUNDARY VALUE PROBLEMS ON GENERAL REGIONS

Klaus Böhmer

Mathematics Research Center
University of Wisconsin—Madison
610 Walnut Street
Madison, Wisconsin 53706

February 1980

(Received September 19, 1978)

(see 1473)

Sponsored by

U.S. Army Research Office
P.O. Box 12211
Research Triangle Park
North Carolina 27709

Deutsche Forschungsgemeinschaft
Bonn, Germany

University of Karlsruhe
Karlsruhe, Germany

UNIVERSITY OF WISCONSIN -MADISON
MATHEMATICS RESEARCH CENTER

HIGH ORDER DIFFERENCE METHODS FOR QUASILINEAR ELLIPTIC BOUNDARY VALUE PROBLEMS ON GENERAL REGIONS

Klaus Böhmer

Technical Summary Report #2042
February 1980

ABSTRACT

Stability and convergence for a difference method for quasilinear elliptic boundary value problems are proved. Asymptotic expansions of the discretization error, basic for Richardson extrapolation, are established. The general theory of "discrete Newton methods" and "iterated defect corrections via neighboring problems" [6,8] and Pereyra's deferred corrections [22] are used to derive different high order methods. Some special cases and computational problems are pointed out and numerical tests are included.

AMS(MOS) Subject Classification - 65N05, 65N10, 65N20, 65B05, 80.35, 80.65

Key Words - elliptic differential equations, difference equations, asymptotic error expansion, defect corrections, stability, heat transfer

Work Unit Number 7 - Numerical Analysis

Sponsored by the United States Army under Contract No. DAAG29-75-C-0024, the Deutsche Forschungsgemeinschaft under Contract No. BO 622/1 and the University of Karlsruhe.

SIGNIFICANCE AND EXPLANATION

In this paper we discuss the numerical solution of quasilinear elliptic boundary value problems in a bounded region Ω of $\mathbb{R}^n$.

Problems of this type arise in many physical and chemical applications, e.g. in chemical reaction processes, in vortex motions of fluids, in steady state heat conduction, in diffusion processes and in structural mechanics.

By applying the usual five point discretization in the interior of Ω and high order approximation of the boundary conditions we obtain stable discretizations with an error admitting an asymptotic expansion. This expansion may be employed with Richardson extrapolation or defect or deferred corrections to obtain methods of flexible order (2, 4 and 5.5).

Due to the increasing complexity of the problems, linear equations often are not adequate, so nonlinear problems have to be discussed*. Error asymptotic results have been obtained before (Pereyra [23], 1970) for those problems but under strong restrictions on the boundary of Ω which essentially are satisfied for rectangular domains. For nonrectangular smooth boundaries of Ω only the Dirichlet problem for the Poisson equation was treated before (Pereyra-Proskurowski-Widlund [24], 1977). This paper gives the first approach for smooth boundaries and nonlinear problems, generalizing the results in [23,24]. Further, our high order methods, based on defect and deferred corrections, save a significant amount of computer time and may be used for adaptive techniques.

*Since only a few special treatments, known in mathematical physics, apply to nonlinear problems, one has to treat most of them by numerical methods.

The responsibility for the wording and views expressed in this descriptive summary lies with MRC, and not with the author of this report.

HIGH ORDER DIFFERENCE METHODS FOR QUASILINEAR ELLIPTIC BOUNDARY VALUE PROBLEMS ON GENERAL REGIONS

Klaus Böhmer

Introduction.

In this paper we discuss the numerical solution of quasilinear elliptic differential equations on a general open bounded region $\Omega \subseteq \mathbb{R}^n$ with function values prescribed along the boundary $\partial\Omega$ (Dirichlet problem). Equations of this type arise in several practical problems, one of them discussed in §1. Since we are mainly interested in high order methods the situation has to be smooth enough.

We apply the usual symmetric five point formulas in all regular mesh points x (all neighboring mesh points of x are in $\bar{\Omega}$). In irregular mesh points (at least one of the mesh neighbors lies outside of $\bar{\Omega}$) we introduce the boundary conditions by using interpolating polynomials of degree $k \leq 6$.

For $k = 0$ and 1 we have the well known first and second order methods due to Gerschgorin [20] and Collatz [14] for linear elliptic equations. For $k \geq 2$ the method is due to Kreiss presented by Pereyra-Proskurowski-Widlund [24] for the Poisson equation.

Now, Wasow [28] has shown by some examples, that an asymptotic expansion of the discretization error is not available if the boundary values are not reproduced exactly enough. By a modification of the five point formula in irregular mesh points Bramble-Hubbard [11] obtain a first term in the asymptotic expansion.

Higher order asymptotic expansions are only known in two special cases: For quasilinear operators Ω has to be such that there are only regular mesh points (Pereyra [23]). This regularity condition is satisfied essentially only for rectangular domains. For general regions Ω asymptotic expansions are known only in the Dirichlet problem for the Poisson equation (Pereyra-Proskurowski-Widlund [24]). In this paper we generalize the asymptotic results of Pereyra-Proskurowski-Widlund to quasilinear Dirichlet problems on general regions. Since we are mainly interested

Sponsored by the United States Army under Contract No. DAAG29-75-C-0024, the Deutsche Forschungsgemeinschaft under Contract No. BO 622/1 and the University of Karlsruhe.

in high order results, the situation has to be smooth enough. Our results, applying Wasow's [28], are obtained here by combining stability and local discretization errors. In [10] we use special techniques to give for linear elliptic equations for $k \leq 4$ considerably better results. Especially the conditions which we have to impose on the coefficients are much less stringent than here and the error asymptotic for $k=4$ is nearly as good as here for $k=6$.

In §2 we formulate the discretization scheme and describe the equations obtained. §3 is devoted to a stability proof for our method for $k \leq 6$ using some of the Lemmas of Kreiss, given in [24]. Since we have to impose in §3 relatively strong conditions on the differential operator obtaining only stability with respect to the ℓ_2-norm, we give another stability proof in §4 under much weaker conditions for the ℓ_∞-norm, but only for $k \leq 4$. For $k > 6$ we can't prove stability.

The convergence of the basic second order method and the existence of an asymptotic expansion of the (global) discretization error are proved in §5 for $k \leq 6$. Two special cases which allow an extended asymptotic expansion are discussed in §6, one of them treated already in connection with iterated deferred corrections by Pereyra [23].

§7 uses the asymptotic results in §§5 and 6 to obtain high order methods via Richardson extrapolation. Further it combines the asymptotic expansion with the general theory of "discrete Newton methods" [8] again providing high order methods. In addition we indicate high order methods based on iterated defect corrections via neighbouring problems [29, 30, 27, 15, 16, 7, 8] and on iterated deferred corrections formalized by Pereyra [22]. The mutual advantages and disadvantages of defect corrections and Richardson extrapolation are discussed. Especially, when all mesh points are regular the asymptotic expansion may be extended, depending essentially on the smoothness of the solution. In that case Richardson extrapolation is not able to compete with the other methods, especially with the discrete Newton methods.

The discussion of several problems arising in the computation of the different methods are the topic of §8. The applicability of fast Laplace solvers (see Pereyra-Proskurowski-Widlund [24] and of Newton methods for the original nonlinear discrete problem are discussed.

Richardson extrapolation, deferred and defect corrections and discrete Newton methods are used in §8 to obtain high order methods.Numerical examples are given in §9.

1. A quasilinear elliptic boundary value problem in chemical physics

In a chemical reaction let T be the temperature, k the thermal conductivity, Q the monomolecular heat of reaction and V the velocity of the reaction. Then the equilibrium of the heat transfer between the heat produced by the chemical reaction and the heat conducted away is characterized in $\Omega \subset \mathbb{R}^n$ by the equation (see Ames [2], p. 180):

$$k\Delta T = -QV \text{ in } \Omega \qquad \left(\Delta := \sum_{i=1}^{n} \frac{\partial^2}{\partial x_i^2}\right). \tag{1.1}$$

Now, V and T are related by the Arrhenius relation, which reads, in the simple case we have assumed here, as

$$V = c\nu \exp\left(-\frac{E}{RT}\right), \tag{1.2}$$

where c is the concentration, ν a scaling factor which may be positive or negative, E the energy of activation, and R the universal gas constant. So we obtain from (1.1) and (1.2) the equation

$$\begin{cases} k\Delta T = -c\nu Q \exp\left(-\frac{E}{RT}\right) \text{ in } \Omega \\ \text{with suitable boundary conditions on } \partial\Omega . \end{cases} \tag{1.3}$$

For constant Q and k, and T close to T_0, (1.3) may be transformed into

$$\begin{cases} \Delta\psi + \mu \exp \psi = 0 \\ \text{with } \mu := \frac{c\cdot\nu\cdot Q\cdot E}{kRT_0^2} \cdot \exp\left(-\frac{E}{RT_0}\right), \ \psi = \frac{E}{RT_o^2}(T-T_o). \end{cases} \tag{1.4}$$

It is possible to give the general solution for (1.4) in the two dimensional case ($n = 2$ in (1.1)), see Ames [2], p. 182. The approach given there may be generalized to the equation (1.3). To obtain a solution satisfying the boundary conditions in (1.3) *one* has still to solve the difficult problem to choose three suitable functions F,f,g such that ψ, defined by $\exp \psi = F(f,g)$, satisfies the boundary conditions (see [2], p. [illegible]).

For $n \geq 3$ or for anisotropic heat conduction this approach is no longer possible and we have to find other methods. Here we are concerned with the corresponding difference analoga to (1.3) and their generalizations and with improvements of the approximate solutions.

2. The basic difference scheme for elliptic problems.

Let a_i, $i = 1,\dots,n$, f and g be continuous real valued functions, defined on an open, bounded and connected $\Omega \subseteq \mathbb{R}^n$, $\Omega \times B$ and $\partial\Omega$ respectively with an appropriate $B \subseteq \mathbb{R}^{n+1}$. Then we define an elliptic differential operator F as $(a_i(\cdot) \geq \underline{a} > 0$, ∇y = [illegible]

$$(2.1)\qquad F : \begin{cases} D := \{y \in C^2(\Omega) \cap C(\bar{\Omega}) \mid \underset{x\in\Omega}{\forall}\ (y(x),\nabla y(x)) \in B\} \to C(\Omega) \times C(\partial\Omega)\ , \\ y \to Fy := \begin{Bmatrix} -\sum\limits_{i=1}^{n} a_i(\cdot) y_{x_i x_i}(\cdot) + f(\cdot,y(\cdot),\nabla y(\cdot)) & \text{in } \Omega \\ y(\cdot) - g(\cdot) \quad \text{in } \partial\Omega \end{Bmatrix}, \\ a_i(\cdot) \geq \underline{a} > 0\ . \end{cases}$$

Now let $Fy = 0$ have a unique solution z in D_z with

$$(2.2)\qquad \begin{cases} D_z := \{y \in D \mid \|y - z\|_{L_\infty(\bar{\Omega})} \leq \rho\}\ , \\ z \text{ is unique solution in } D_z \text{ of} \\ 0 = Fz = \begin{Bmatrix} -\sum\limits_{i=1}^{n} a_i(\cdot) z_{x_i x_i}(\cdot) + f(\cdot,z(\cdot),\nabla z(\cdot)) = 0 & \text{in } \Omega \\ z(\cdot) - g(\cdot) = 0 \quad \text{in } \partial\Omega \end{Bmatrix}. \end{cases}$$

Conditions for unique solutions are given, e.g., in Bers [3].

To define a discretization we introduce a grid $\Gamma_{h,n}$ and the grid lines $G_{h,\nu}$ as

$$(2.3)\qquad \begin{cases} \Gamma_{h,n} : = \{x \in \mathbb{R}^n \mid x = (x_1,\dots,x_n)^T, x_i = n_i h, n_i \in \mathbb{Z}\}\ , \\ G_{h,n} : = \{x \in \mathbb{R}^n \mid x = (x_1,\dots,x_n)^T, x_\nu \in \mathbb{R}, x_i = n_i h, n_i \in \mathbb{Z} \text{ for } i \neq \nu, \nu = 1,\dots,n\}, \end{cases}$$

where we have, for simplicity, chosen an equal stepsize h in all directions. The following discretization is a generalization of methods given by Gerschgorin [2] and Collatz [14] for linear elliptic second order equations $(k = 0$ and $k = 1)$ and by Kreiss and Pereyra-Proskurowski-Widlund [24] $(k \leq 6)$ for Poisson's equation.

With the open set Ω we introduce the mesh points in Ω and distinguish between the set of regular mesh points (slightly different from the Introduction)

$$\Omega_h := \{x \in \Omega \mid x \pm h e_i \in \Omega,\ i = 1,\dots,n\} \cap \Gamma_{h,n}$$

(here e_i are the unit-vectors in the direction of the positive i-th coordinate axis) and the set of irregular (mesh) points

$$\Omega_{h,i} := \{x \in \Omega_h | \exists i : x + he_i \notin \bar\Omega \text{ or } x - he_i \notin \bar\Omega\} \cap \Omega_{h,n} .$$

In regular points we use the standard centered difference approximations for the derivatives to obtain the [illegible] in the form

$$(2.4) \quad \begin{cases} h^2(\Phi_h F)\eta_h(x) := -\sum_{i=1}^{n} a_i(x)\{\eta_h(x + he_i) - 2\eta_h(x) + \eta_h(x - he_i)\} \\ \quad + h^2 f\left(x, \eta_h(x), \dfrac{\eta_h(x + he_1) - \eta_h(x - he_1)}{2h}, \dots, \dfrac{\eta_h(x + he_n) - \eta_h(x - he_n)}{2h}\right) \\ \quad \text{for } x \in \Omega_h . \end{cases}$$

If x is an irregular mesh point, at least one of the $x \pm he_i \notin \bar\Omega$. In this case we have to replace in (2.4) every $\eta_h(x \pm he_i)$ with $x \pm he_i \notin \bar\Omega$ by a provisional value obtained by polynomial extrapolation. Let, for that purpose, $x \in \Omega_{h,i}$, $x + he_i \notin \bar\Omega$, but $x - he_i, \dots, x - (k - 1)he_i \in \Omega_h$. Further let x_i^*, $x < x_i^* \leq x + he_i$, be the unique intersection of $\partial\Omega$ and the line segment $[x, x + he_i]$. For h small and $\partial\Omega$ smooth enough, these conditions will be satisfied.

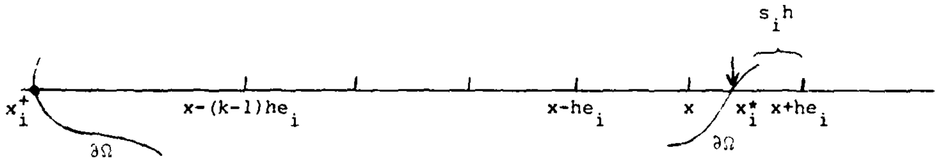


Figure 1

Now we define the following approximations for $y_{x_ix_i}$ (resp. y_{x_i}): Compute an interpolating polynomial P_k of degree k, defined by $P_k(x - (\nu - 1)he_i) = y_{1-\nu}$, $\nu = 1,\dots,k$, $P_k(x_i^*) = g(x_i^*)$, and replace in $h^2y_{x_ix_i} \approx y_1 - 2y_0 + y_{-1}$ resp. $2hy_{x_i} \approx y_1 - y_{-1}$ the value y_1 by $y_1 := P_k(x + he_i)$. If the distance from x_i^* to $x + he_i$ is s_ih, with $0 \leq s_i < 1$, we find (see Pereyra-Proskurowski-Widlund [24])

$$(2.5)\quad \begin{cases} y_1 = \dfrac{1}{\alpha_{0,i}}\left(y(x_i^*) - \sum\limits_{\nu=1}^{k} \alpha_{\nu,i}\, y_{1-\nu}\right) \\[2ex] \text{with } \alpha_\nu := \alpha_{\nu,i} := \prod\limits_{\substack{\ell=0\\ \ell\neq\nu}}^{k} (s_i - \ell)/(\nu - \ell), \qquad i = 1,\dots,n \end{cases}$$

If $s_i = 0$, that is $x_i^* = x + he_i$ we find $\alpha_{0,i} = 1$, $\alpha_{\nu,i} = 0$, $\nu = 1,\dots,k$. Therefore $y_1 = y(x_i^*)$, and the following equation (2.6) reduces to (2.4) if $s_i = 0$, $i = \mu_1,\dots,\mu_m$ and if we use $\eta_h(x + he_{\mu_i}) = g(x + he_{\mu_i})$ in (2.4).

For formal reasons we take h so small that with $x \pm he_i \notin \Omega_h$ we have $x \mp (\nu-1)he_i \in \Omega_h$ for $\nu=1,\dots,k-1$. Now we formulate the irregular part of the equations corresponding to (2.4) for the case that only $x + he_{\mu_1},\dots,x+he_{\mu_m} \notin \Omega$ and no $x - he_i \notin \Omega$. We obtain then for $x=(x_1,\dots,x_n)$ and $g(x_i^*):=g(x_1,\dots,x_{i-1},x_i^*,x_{i+1},\dots,x_n)$

$$(2.6)\quad \left\{ \begin{aligned} h^2(\varphi_h F)\eta_h(x) := &- \sum_{\substack{i=1\\ i\neq\mu_1,\dots,\mu_m}}^{n} a_i(x)\{\eta_h(x + he_i) - 2\eta_h(x) + \eta_h(x - he_i)\} \\ &- \sum_{i=\mu_1,\dots,\mu_m} a_i(x)\left\{\frac{1}{\alpha_{0,i}}\{g(x_i^*) - \sum_{\nu=1}^{k} \alpha_{\nu,i}\eta_h(x - (\nu-1)he_i)\} - 2\eta_h(x)\right. \\ &\left. + \eta_h(x - he_i)\right\} + h^2 f(x,\eta_h(x), \frac{\eta_h(x + he_1) - \eta_h(x - he_1)}{2h},\dots, \\ &\frac{\frac{1}{\alpha_{0,\mu_1}}\{g(x_{\mu_1}^*) - \sum_{\nu=1}^{k} \alpha_{\nu,\mu_1}\eta_h(x - (\nu-1)he_{\mu_1}))\} - \eta_h(x - he_{\mu_1})}{2h},\dots \\ &\frac{\frac{1}{\alpha_{0,\mu_m}}\{g(x_{\mu_m}^*) - \sum_{\nu=1}^{k} \alpha_{\nu,\mu_m}\eta_h(x - (\nu-1)he_{\mu_m})\} - \eta_h(x - he_{\mu_m})}{2h},\dots \\ &\dots, \frac{\eta_h(x + he_n) - \eta_h(x - he_n)}{2h}), \quad \text{for } x \in \Omega_{h,i}\,. \end{aligned} \right.$$

If instead of $x + he_\mu \notin \Omega$ we have $x - he_\mu \notin \Omega$, it is obvious how the equations have to be changed. We finally add the boundary values explicitely by

$$(2.7)\qquad (\varphi_h F)\eta_h(x) := \eta_h(x) - g(x) \text{ for } x \in G_{h,n} \cap \partial\Omega.$$

When stability and consistency are proved then (1.1) implies for h small [illegible], see Stetter [26], the unique solvability of $(\varphi_h F)\eta_h = $, that is

$$(2.8)\quad \begin{cases} \Delta_h : \begin{cases} E := C^2(\Omega) \cap C(\bar{\Omega}) \rightarrow E_h := \{\eta_h : (\Gamma_{h,n} \cap \Omega) \cup (G_{h,n} \cap \partial\Omega) \rightarrow \mathbb{R}\} \\ y \rightarrow y|(\Gamma_{h,n} \cap \partial\Omega) \cup (G_{h,n} \cap \partial\Omega) \end{cases} \\ \zeta_h \in D_{h,z} := \Delta_h D_z \text{ is the unique solution of } (\varphi_h F)\eta_h = \text{ for small } h. \end{cases}$$

Again by, e.g. Stetter [26], we need for the proof of the stability of the [illegible] $(\varphi_h F)$ only to discuss the stability of $(\varphi_h F)'(\eta_h)$ for $\|\eta_h - \Delta_h z\|$ small enough. [illegible] (2.4) and (2.6) imply that, for $f \in C^{1+\alpha}(\Omega)$ and $y \in C^{\prime}(\bar{\Omega})$ with essentially [illegible] derivatives, $\varphi_h F'(y) = (\varphi_h F)'(\Delta_h y) + O(h^2)$. That means, we may restrict our stability proof to the discussions of $\varphi_h F'(z_o)$: So we formulate the discretization φ_h [illegible] affine operator of the form (2.1). With continuous real valued functions $a_i(\cdot)$, $b_i(\cdot)$, $i = 1,\ldots,n$, $c(\cdot)$ and $d(\cdot)$, defined on Ω, $a_i(\cdot) \geq \underline{a} > 0$, and $g(\cdot)$ defined on $\partial\Omega$ we have

$$(2.9)\quad F_d \begin{cases} D_0 := C^2(\Omega) \cap C(\bar{\Omega}) \rightarrow C(\Omega) \times C(\bar{\Omega}) \\ y \rightarrow F_d y := \begin{Bmatrix} \sum\limits_{i=1}^{n} (-a_i(\cdot)y_{x_i x_i} + b_i(\cdot)y_{x_i}) + c(\cdot)y(\cdot) + d(\cdot) \\ \text{in } \Omega, \quad y(\cdot) - g(\cdot) \text{ in } \partial\Omega \end{Bmatrix}. \end{cases}$$

In regular points the discretization $\varphi_h F_d$ in (2.4) reduces here to

$$(2.10)\quad \begin{cases} h^2(\varphi_h F_d)\eta_h(x) = \left(2 \sum\limits_{i=1}^{n} a_i(x) + h^2 c(x)\right)\eta_h(x) \\ + \sum\limits_{i=1}^{n} \{-a_i(x) + \frac{h}{2} b_i(x)\}\eta_h(x + he_i) + \sum\limits_{i=1}^{n} \{-a_i(x) - \frac{h}{2} b_i(x)\}\eta_h(x - he_i) \\ + h^2 d(x) \qquad \text{for } x \in \Omega_h . \end{cases}$$

In irregular points those $\eta_h(x \pm he_i)$ in (2.10) have to be replaced by expressions corresponding to (2.5) for which $x \pm he_i \notin \Omega$. For the stability proof in §3 it is only important that $\varphi_h F_0$, $F_0 := F_d|_{d=0}$, may be separated into the sum of discretizations for ordinary second order equations. So we have for regular points:

$$
(2.11) \quad \begin{cases} h^2(\varphi_h F_0)\eta_h(x) = \sum_{i=1}^{n} a_i(x)\left\{\left[2 + h^2 \frac{c(x)}{\sum_{\nu=1}^{n} a_\nu(x)}\right]\eta_h(x) + \right. \\[2ex] \left. + \left[-1 + \frac{h}{2}\frac{b_i(x)}{a_i(x)}\right]\eta_h(x+he_i) + \left[-1 - \frac{h}{2}\frac{b_i(x)}{a_i(x)}\right]\eta_h(x-he_i)\right\} \quad \text{for } x \in \text{[illegible]} . \end{cases}
$$

Again, for irregular points the $\eta_h(x \pm he_i)$ in (2.11) with $x \pm he_i \notin$ [illegible] have to be replaced by expressions corresponding to (2.5). Further, we add (2.7) for the boundary points to obtain a system of linear equations. [illegible] it is enough to study the properties of the matrix A defined by

$$
(2.12) \qquad h^2(\varphi_h F_0)\eta_h(x) \quad \text{for } x \in \Omega_h \text{ and for } x \in \Gamma_{h,i}
$$

This matrix A may be written in the form (see Pereyra-Proskurowski-Widlund [24])

$$
(2.13) \qquad A = \sum_{i=1}^{n} P_i^T A_i P_i
$$

with suitable permutation matrices P_i. The matrices A_i are obtained by collecting only those contributions in $(\varphi_h F_0)\eta_h(x)$, which are multiplied by a fixed $a_i(x)$ (see (2.11)). Therefore these A_i are direct sums of matrices of the form (2.14), where we use the abbreviations to be introduced in (2.15)

(2.14) B:=

$$
\begin{array}{l} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_{\ell-1} \\ a_\ell \end{array}
\left(\begin{array}{llllllll}
2+h^2 q_0+\frac{\alpha_1'}{\alpha_0'}p_0^+, & -p_0^-+\frac{\alpha_2'}{\alpha_0'}p_0^+, & \frac{\alpha_3'}{\alpha_0'}p_0^+, & \cdot, \cdot, \ldots, & \frac{\alpha_k'}{\alpha_0'}p_0^+, & 0, & \ldots\ldots, & \text{[illegible]} \\
-p_1^+, & 2+h^2 q_1, & -p_1^-, & 0, 0, \ldots, & 0, & 0, & \ldots\ldots, & \\
0, & -p_2^+, & 2+h^2 q_2, -p_2^-, & 0, \ldots, & 0, & 0, & \ldots\ldots, & \text{[illegible]} \\
\vdots & & & & & & & \\
0, & \ldots\ldots\ldots & & , 0, \ldots\ldots, & 0, & -p_{\ell-1}^+, & 2+h^2 q_{\ell-1}, & -p_{\ell-1}^- \\
0, & \ldots\ldots\ldots & & , \frac{\alpha_k}{\alpha_0}p_\ell^+, \ldots\ldots, & \frac{\alpha_3}{\alpha_0}p_\ell^+, & -p_\ell^+ + \frac{\alpha_2}{\alpha_0}p_\ell^+, & 2+h^2 q_\ell+\frac{\alpha_1}{\alpha_0}p_\ell^+ &
\end{array}\right)
$$

The $a_\nu = a_\nu(x)$, see([illegible]), in front of the matrix give the factors with which [illegible] has to be multiplied. Each matrix B corresponds to a maximal intersecting interval $[x_i^+, x_i^*] \subset \bar{\Omega} \cap G_{h,n}$ in the direction of the coordinate vector e_i, that is $x_i^+, x_i^* \in \partial\Omega$, see Figure 2. For simplicity we have used the following abbreviations

$$
(2.15)\quad \begin{cases}
a(\cdot) := a_i(\cdot)\Big|_{[x_i^+,x_i^*]}, \; p(\cdot) := \dfrac{b_i(\cdot)}{a_i(\cdot)}\Big|_{[x_i^+,x_i^*]}, \; q(\cdot) := \dfrac{c(\cdot)}{\sum\limits_{\nu=1}^{n} a_\nu(\cdot)}\Big|_{[x_i^+,x_i^*]} \\
\alpha_\nu := \alpha_{\nu,i}, \alpha'_\nu := \alpha'_{\nu,i} \text{ where } \alpha'_{\nu,i} \text{ correspond to the first} \\
\text{intersection } x_i^+ \text{ of } \partial\Omega \text{ and } \{x - te_i \mid t > 0\} \text{ left of } x_i^*. \text{ Further,} \\
\text{let } a_\nu := a(x_\nu),\; p_\nu := p(x_\nu),\; q_\nu := q(x_\nu), \text{ where } x_\nu \text{ are defined by Figure 2 and} \\
p_\nu^+ := 1 + \frac{h}{2} p_\nu, \; p_\nu^- := 1 - \frac{h}{2} p_\nu.
\end{cases}
$$

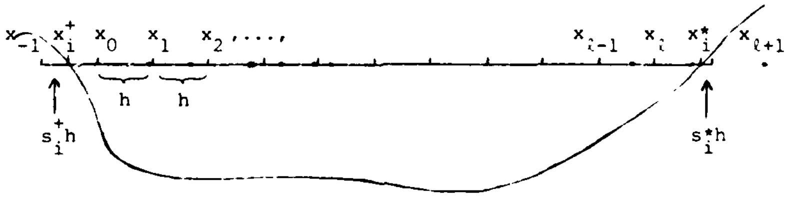


Figure 2

In our discussion we have confined the coefficients $a_i(\cdot)$ to be functions of the independent variable $x \in \Omega$ alone. Yet, we could generalize the whole discussion in §§2 ff. to the case

$$
(2.16)\qquad Fz := \left\{ \begin{array}{l} -\sum\limits_{i=1}^{n} a_i(\cdot, z(\cdot), z_{x_i}(\cdot)) z_{x_i x_i}(\cdot) + f(\cdot, z(\cdot), \nabla z(\cdot)) \text{ in } \Omega \\ z(\cdot) - g(\cdot) \text{ on } \partial\Omega \end{array} \right\} = 0
$$

with $a_i(\cdot, z(\cdot), z_{x_i}(\cdot)) \geq \underline{a} > 0$.

Then we would have the derivative

$$(2.17)\quad F'(y)u = \begin{cases} -\sum_{i=1}^{n} \{a_i(\cdot,y(\cdot),y_{x_i}(\cdot))u_{x_i x_i}(\cdot) + a_i^{(0,1,0)}(\cdot,y(\cdot),y_{x_i}(\cdot))u(\cdot) \\ + a_i^{(0,0,1)}(\cdot,y(\cdot),y_{x_i}(\cdot))u_{x_i} + f^{(0,1,\ldots,0)}(\cdot,y(\cdot),\nabla y(\cdot))u(\cdot) \\ + \sum_{i=1}^{n} f^{(\underbrace{0,\ldots,1}_{i+2},\ldots,0)}(\cdot,y(\cdot),\nabla y(\cdot))u_{x_i}(\cdot) \quad \text{in } \Omega,\ u(\cdot) \text{ on } \partial\Omega \end{cases}$$

which is of the same structure, but more complicated than [illegible].

In §3 we give a stability proof for (2.12) and $k \leq 6$ by showing that $\|A\|_2 \leq Ch^{-2}$ for the matrix A defined in (2.12). So all results based on §3 are results for the Euclidean norm. The conditions which we have to impose are rather stringent. Therefore we give another, much easier, stability proof, valid only for $k \leq 4$ showing that $\|A\|_\infty \leq Ch^{-2}$. So, using this §4 with its less stringent conditions sup-norm results are obtained. Since k essentially regulates the number [illegible] in the asymptotic expansion both considerations are worthwhile.

3. Stability

By stability we mean the uniform boundedness of the inverse operator [illegible] the difference operator in a neighbourhood of the exact [illegible]. To prove this property for the method, defined in (2.4)-(2.7), it is enough to [illegible] stability of the discretization for the linear problem F_z in (2.9), where $F_z y := F'(z)y$ and F from ([illegible]) (see [illegible] [[illegible]], p. [illegible]). , we [illegible] prove that the matrix A, defined by (2.12), has the property $\|A^{-1}\| \leq h^{-2}$. For that purpose we use and generalize some of the Lemmas, due to Kreiss, given in Pereyra-Proskurowski-Widlund [24]. The norms in this paragraph are exclusively the Euclidean vector norm and the spectral matrix norm $(\|\cdot\|_2)$.

Lemma 3.1 [24]: Let the symmetric part of a matrix A satisfy

$$(A + A^T)/2 \geq \delta I, \qquad \delta > 0 .$$

Then A is regular and $\|A^{-1}\| \leq \delta^{-1}$.

Lemma 3.2 [24]: Let $A = \sum_{i=1}^{n} P_i^T A_i P_i$ with permutation matrices P_i. If

$$(A_i + A_i^T)/2 \geq \delta I, \qquad \delta > 0 \text{, for all } i.$$

Then

$$(A + A^T)/2 \geq n\delta I .$$

Lemma 3.3 [24]: Let the matrix A_i be the direct sum of certain matrices B_{ij}. If

(3.1) $$(B_{ij} + B_{ij}^T)/2 \geq \delta I, \qquad \delta > 0, \text{ for all } j ,$$

Then

$$(A_i + A_i^T)/2 \geq \delta I .$$

We have seen in §2 that the matrix A in (2.13) is the sum of matrices $P_i^T A_i P_i$, where the A_i are direct sums of matrices of the form (2.14) and P_i are permutation matrices. we have assumed h so small that $x \in \Omega_{h,i}$ implies, e.g., $x + he_i \notin \Omega$ and $x - (\nu - 1)he_i \in \Omega_h$ for $\nu = 1,\dots,2k - 1$. Especially we have therefore in (2.14) that $\ell \geq 2k - 1$. With the functions a, p, q, and $a_\nu, \varphi_\nu, q_\nu$ introduced in ([illegible]), [illegible] with indices changed for convenience, we may prove as in [24] the following

Lemma 3.4: Let the matrix B in (2.14) be split into two matrices B_1, B_2, where, with p_ν^+ and p_ν^- in (2.15),

(3.2) $B_1 :=$

$$\begin{array}{c|ccccccc}
a_1 & 1+\frac{h^2}{2}q_1 & -p_1^- & 0 & & & & \bigcirc \\
a_2 & -p_2^+ & 2+h^2q_2 & -p_2^- & \ddots & & & \\
 & & \ddots & \ddots & \ddots & \ddots & & \\
 & \bigcirc & & \ddots & \ddots & \ddots & \ddots & \\
a_{m-1} & 0 & 0 & \ldots\ldots\ldots\ldots , 0, & -p_{m-1}^+ , & 2+h^2q_{m-1} , & -p_{m-1}^- & \\
a_m & 0 & 0 & \ldots\ldots\ 0,\ p_{m\alpha_o}^{+\alpha_k} , \ldots , & p_{m\alpha_o}^{+\alpha_3} , & -p_m^+ + p_{m\alpha_o}^{+\alpha_2} , & 2+h^2q_m+p_{m\alpha_o}^{+\alpha_1} &
\end{array}$$

Let B_2 be obtained from B_1 by inverting the order of rows and columns and changing m to m' and α_ν to α_ν'. Further let

$$m + m' = \ell + 1, \quad m \geq k, \quad m' \geq k .$$

If

$$(B_1 + B_1^T)/2 \geq \delta I \quad \text{and} \quad (B_2 + B_2^T)/2 \geq \delta I ,$$

then

$$(B + B^T)/2 \geq \delta I .$$

Before we are able to prove the next Lemma we have to do some preparations. We obtain the symmetric part S of B_1 as

$$(3.3) \qquad S := (B_1 + B_1^T)/2 = \begin{pmatrix}
s_{11} & s_{21} & & & & & 0 \\
s_{21} & s_{22} & s_{32} & & \bigcirc & & \vdots \\
 & s_{32} & s_{33} & s_{34} & & & 0 \\
 & & \ddots & \ddots & \ddots & & s_{mm-k+1} \\
 & \bigcirc & & \ddots & \ddots & \ddots & \vdots \\
 & & & & s_{m-2m-1} & s_{m-1m-1} & s_{mm-1} \\
0 & \cdots & 0 & s_{mm-k+1} & \cdots & s_{mm-1} & s_{mm}
\end{pmatrix}$$

where

$$\left\{\begin{array}{ll}
s_{11} = a_1(1 + \frac{h^2}{2} q_1), \quad s_{\nu\nu} = a_\nu(2 + h^2 q_\nu), & \nu = 2,\ldots,m-1, \\
s_{mm} = a_m\left(2 + h^2 q_m + (1 + \frac{h}{2} p_m) \frac{\alpha_1}{\alpha_0}\right), & \\
s_{\nu+1\nu} = -\left(\frac{a_{\nu+1} + a_\nu}{2} + \frac{h}{4}(a_{\nu+1}p_{\nu+1} - a_\nu p_\nu)\right), & \nu = 2,\ldots,m-2, \\
s_{m\nu} = +a_m(1 + \frac{h}{2} p_m) \frac{\alpha_{m-\nu+1}}{2\alpha_0}, & \nu = m-k+1,\ldots,m-1, \\
s_{mm-1} = -\left(\frac{a_m + a_{m-1}}{2} + \frac{h}{4}(a_m p_m - a_{m-1}p_{m-1}) - a_m(1 + \frac{h}{2} p_m) \frac{\alpha_2}{2\alpha_0}\right), & \\
s_{\nu\mu} = 0 \text{ elsewhere} . &
\end{array}\right. \tag{3.4}$$

Let us now assume that

$$\left\{\begin{array}{l}
a \in C^{3+\alpha}[x_i^+, x_i^*], \quad p \in C^{2+\alpha}[x_i^+, x_i^*], \quad q \in C^{1+\alpha}[x_i^+, x_i^*], \qquad 0 < \alpha < 1, \\
a(\cdot) \geq \underline{a} > 0 \text{ in } [x_i^+, x_i^*]. \text{ There exists} \\
\hat{x} \in [x_i^+ + (k+1 - s_i^+)h,\ x_i^* - (k + 1 - s_i^*)h] \text{ with} \\
a'(\hat{x}) = 0, \quad a'(\cdot) \geq 0 \text{ in } [x_i^+, \hat{x}] \text{ and } a'(\cdot) \leq 0 \text{ in } [\hat{x}, x_i^*]. \\
\left(2aq - a'' + (\operatorname{sgn} a')(ap)' + \frac{a'^2}{a}\right)(\cdot) \geq c_+ > 0 \text{ in } [x_i^+, x_i^*], \\
\left(2aq + (\operatorname{sgn} a')(ap)' - \frac{a'^2}{2a}\right)(\cdot) \geq 0 \text{ in } [x_i^+, x_i^*], \text{ where we define} \\
\{2aq - |(ap)'|\}(\cdot) \geq c_+ > 0 \text{ in an interval of length} \geq h \text{ with the midpoint } \hat{x}. \\
\operatorname{sgn} a' := \left\{\begin{array}{ll} 1 \text{ in } [x_i^+, \hat{x}) & \text{for } a \geq 0 \\ -1 \text{ in } (\hat{x}, x_i^*] & \text{for } a' < 0 . \end{array}\right.
\end{array}\right. \tag{3.5}$$

Here $C^{m+\alpha}[x_i^+, x_i^*]$ means that all derivatives up to the order m are (uniformly) Hölder-continuous in $[x_i^+, x_i^*]$ with exponent α, $0 < \alpha < 1$. For $\left(\frac{3}{2} \cdot \frac{a'^2}{a} - a''\right)(\cdot) \leq c_+$ the forelast inequality ($\geq c+$) in (3.5) includes the last (≥ 0).We will come back to the case $a'(\cdot) \leq 0$ in $[x_i^+, \hat{x}]$, $a'(\cdot) \geq 0$ in $[\hat{x}, x_i^*]$ in Remark 3.7.

Now we choose the numbering in (3.2) such that $|\hat{x} - x_1| \leq \frac{h}{2}$.

Assumptions (3.5) imply that for sufficiently small h we have $s_{11} \geq \underline{a} > 0$, $s_{\nu\nu} \geq \underline{a}$, $\nu=1,\dots,m$ and $s_{\nu+1\nu} \leq 0$, $\nu=1,\dots,m-1$. To show that S is positive definite we use its LDL^T representation with a diagonal matrix D. We first give the L-R factorization as

$$(3.6)\qquad S = L \cdot R = \begin{pmatrix} 1 & & & & \bigcirc \\ \ell_{21} & 1 & & & \\ & \ell_{32} & 1 & & \\ & & \ddots & \ddots & \\ \bigcirc & & & \ddots & \ddots \\ 0 & \dots & 0 & \ell_{m\,m-k+1} & \dots & 1 \end{pmatrix} \begin{pmatrix} r_1 & s_{21} & & & \cdot \\ & r_2 & s_{32} & \bigcirc & \cdot \\ & & \ddots & \ddots & s_{m,m-k+1} \\ & \bigcirc & & \ddots & \cdot \\ & & & r_{m-1} & s_{m\,m-1} \\ & & & & r_m \end{pmatrix}$$

where the $\ell_{\mu\nu}$ and the r_ν are defined in the usual way, so, e.g.,

$$(3.7)\qquad \begin{cases} r_1 := s_{11}, \quad \ell_{\nu+1\nu} := s_{\nu+1\nu}/r_\nu, & \nu = 1,\dots,m-2 \\ r_{\nu+1} := -\ell_{\nu+1\nu} \cdot s_{\nu+1\nu} + s_{\nu+1\nu+1}, & \nu = 1,\dots,m-2\,. \end{cases}$$

To give the $\ell_{m\nu}$ and r_m we would have to introduce the whole algorithm. Since we need only the properties, given in (3.13) for the $\ell_{m\nu}$ and in (3.14) for r_m we do not give the explicit results corresponding to (3.7).

Using the symmetry of S and the properties of the $\ell_{\nu+1\nu}$ one straight foreward verifies

$$(3.8)\qquad S = L \cdot \begin{pmatrix} r_1 & & & \bigcirc \\ & \ddots & & \\ & & r_{m-1} & \\ \bigcirc & & & r_m \end{pmatrix} \cdot L^T = LDL^T\,.$$

So we will study the matrices

$$(3.9)\qquad L = \begin{pmatrix} L_{11} & 0 \\ \ell^T & 1 \end{pmatrix} \quad \text{and} \quad D = \begin{pmatrix} r_1 & & & \bigcirc \\ & r_2 & & \\ & & \ddots & \\ & & & r_{m-1} & \\ \bigcirc & & & & r_m \end{pmatrix}.$$

We obtain from (3.4)

$$(3.10)\qquad s_{\nu+1\nu} = -\{a + \frac{h}{2}a' + \frac{h^2}{4}[(ap)' + a'']\}_\nu + O(h^3)$$

and, by induction, one shows that for $\nu = 1,\dots,m-1$

$$(3.11)\qquad r_\nu = a_\nu + \frac{h^2}{4}\left\{\sum_{\mu=1}^{\nu-1} 4aq - 2(ap)' - \frac{a'^2}{a} \text{ [illegible] } - 2(aq)_\nu + 4(aq)_\nu\right\} + O(h^3) .$$

For the inductive proof and for our later discussions we need $\ell_{\nu+1\nu}$ and we find by (3.7), (3.10), (3.11) that

$$(3.12)\qquad \begin{cases} \ell_{\nu+1\nu} = -1 - \frac{h}{2}\frac{a_\nu'}{a_\nu} - \frac{h^2}{4a_\nu}\left\{\sum_{\mu=1}^{\nu-1}\left(2(ap)' + \frac{a'^2}{a} - 4aq\right)_\mu + (ap)'_\nu + \text{[illegible]}\right. \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \left. -(aq)_\nu + a''_\nu + \text{[illegible]}\right\} + O(h^3) \\ \text{for}\quad \nu = 1,2,\dots,m-2 \end{cases}$$

The vector $\ell^T = (0,\dots,0,\ell_{mm-k+1},\dots,\ell_{mm-1})$ in (3.9) is proved, by similar arguments, to be of the form

$$(3.13)\quad \ell^T = \frac{1}{2\alpha_0}(0,\dots,0,\alpha_k,\alpha_k + \alpha_{k-1},\dots,\alpha_k + \cdots + \alpha_2) + (0,\dots,0,\underbrace{O(h),\dots,O(h)}_{k-1}) .$$

Further (3.8) and (3.13) imply

$$(3.14)\quad \begin{cases} r_m = s_{mm} - \sum_{\nu=m-k+1}^{m-1} r_\nu \ell^2_{m\nu} = a_m\left\{2 + \frac{\alpha_1}{\alpha_0}\right. \\ \left. - \frac{1}{4\alpha_0^2}[\alpha_k^2 + \cdots + (\alpha_k + \cdots + \alpha_s)^2] - \left[\frac{\alpha_k + \cdots + \alpha_2}{2\alpha_0} - 1\right]^2\right\} + O(h) = d_k(s) + O(h) , \end{cases}$$

where we have used (see [24] and (2.5))

$$(3.15)\qquad d_k(s) = r_m(h = 0) = a_m\left\{\frac{1}{\alpha_0} - \frac{\alpha_k^2 + \cdots + (\alpha_k + \cdots + \alpha_2)^2}{4\alpha_0^2}\right\} .$$

This rational function d_k satisfies (see [24])

$$d_k(s) \geq c_+(k) > 0 \quad \text{for} \quad k = 1,2,\dots,6 \quad \text{and} \quad 0 \leq s \leq 1 ,$$

whereas it changes sign for $k = 7$ and 8. Therefore, for h sufficiently small, we again have

(3.16) $$r_m(s) \geq \frac{c_+(k)}{2} > 0 \quad \text{for} \quad k = 1,\ldots,6,\ h \quad \text{small} .$$

These results allow the generalization of Lemma 5 in Pereyra-Proskurowski-Widlund [24]:

Lemma 3.5: Let d_{min} denote the minimum in $[0,1]$ of the function d_k defined in (3.15), let (3.5) be satisfied and h be small enough. Then there is a positive constant C, independent of the mesh size h and the region Ω, such that

(3.17) $$\begin{cases} S \geq \delta I \quad \text{with} \\ \delta = Cd_{min}h^2/(\operatorname{diam}(\Omega))^2 . \end{cases}$$

Proof: Since we need estimates of the form (3.17) we give lower estimates for

$$x^T Sx = x^T LDL^T x \geq \min_{\nu=1}^{m}\{r_\nu\} \cdot \|L^T x\|_2^2 .$$

Now by (3.5) and with $\nu \leq m \leq O(\frac{1}{h})$ we find by (3.11) $r_\nu = a_\nu + O(h)$, so $r_\nu \geq a_\nu/2 \geq \underline{a}/2$, $\nu = 1,2,\ldots,m-1$, and with (3.16)

(3.18) $$x^T Sx \geq \min\{\underline{a}, c_+(k)\}/2 \cdot \|L^T x\|_2^2 .$$

We try to find an upper bound for $\|L^{-T}y\|_2$ since $\|L^{-T}y\|_2 \leq c\|y\|_2$, $c \in \mathbb{R}_+$, implies $\|L^T x\|_2 \geq c^{-1}\|x\|_2$. With $y_m \in \mathbb{R}$, $y^T = (\tilde{y}^T, y_m)$ and ℓ^T, L_{11} in (3.9), one verifies, since L is regular,

$$y^T L^{-1} = ((\tilde{y}^T - y_m \ell^T)L_{11}^{-1}, y_m) ,$$

simply by multiplying the equation from right with L. Since the α_ν/α_0 are of the form $\beta_{\nu,0}(k) \cdot \frac{s}{\nu - s}$ and since the vector ℓ in (3.13) includes only α_ν/α_0, $\nu = 2,\ldots,k$, ℓ has a uniformly bounded norm in $0 \leq s \leq 1$. So

$$\|L^{-T}y\|_2^2 \leq \|L_{11}^{-1}\|_2^2(\|\tilde{y}^T\|_2 + |y_m|\,\|\ell\|_2)^2 + |y_m|^2 \leq C(\|L_{11}^{-1}\|_2^2 + 1)\|y\|_2^2 .$$

Now, since

$$\|L_{11}^{-1}\|_2^2 = (\text{smallest eigenvalue of } L_{11}L_{11}^T)^{-1} ,$$

we are going to estimate this eigenvalue. The definition of L_{11} in (3.6) and (3.9) leads to

$$L_{11}L_{11}^T = \begin{pmatrix} 1 & \ell_{21} & & & & \\ \ell_{21} & 1+\ell_{21}^2 & \ell_{32} & & \bigcirc & \\ & \ddots & \ddots & \ddots & & \\ & & \ell_{\nu\nu-1} & 1+\ell_{\nu\nu-1}^2 & \ell_{\nu+1\nu} & \\ & \bigcirc & & \ddots & \ddots & \ell_{m-1m-2} \\ & & & & \ell_{m-1m-2} & 1+\ell_{m-1m-2}^2 \end{pmatrix} .$$

We use the well-known theorem of Gerschgorin [18] to prove

$$(3.19) \qquad \|L_{11}^{-1}\|_2^{-2} \geq \min\begin{Bmatrix} 1-|\ell_{21}|,\ 1+\ell_{m-1m-2}^2-|\ell_{m-1m-2}|, \\ 1+\ell_{\nu\nu-1}^2-|\ell_{\nu\nu-1}|-|\ell_{\nu+1\nu}|, \quad \nu=2,\ldots,m-2 \end{Bmatrix} \geq c_1h^2$$

with $c_1 \in \mathbb{R}_+$. We find, with $|x_i-\hat{x}| \leq \frac{h}{2}$, $a''(\hat{x}) \leq 0$ and $\ell_{\nu+1\nu} = -1+O(h)$

$$(3.20) \begin{cases} 1-|\ell_{21}| \geq \frac{h^2}{4a_1}\{2aq-(ap)'\}_1 \text{ for } x_1 \leq \hat{x}, \\ 1-|\ell_{21}| \geq \frac{h^2}{4a_1}\{2aq-(ap)'-a''\}_1 \text{ for } x_1 \geq \hat{x}, \\ 1+\ell_{m-1m-2}^2-|\ell_{m-1m-2}| = 1+O(h). \end{cases}$$

Ignoring, for a moment, the $\nu O(h^3)$-terms in (3.12) we find

$$(3.21) \begin{cases} 1+\ell_{\nu\nu-1}^2-|\ell_{\nu\nu-1}|-|\ell_{\nu+1\nu}| = (1+\ell_{\nu\nu-1})^2+(\ell_{\nu+1\nu}-\ell_{\nu\nu-1}) \\ \geq \frac{h^2}{2}\left\{2q-(\frac{a''}{a})-\frac{(ap)'}{a}\right\}_\nu \\ +\frac{h^3}{2}\frac{a'_{\nu-1}}{a_{\nu-1}^2}\sum_{\mu=1}^{\nu-2}\left(2(ap)'+\frac{a'^2}{a}-4aq\right)_\mu + O(h^3) . \end{cases}$$

Since ν might be $[\frac{1}{h}]$ we cannot neglect the $h^3 \cdot \Sigma$ - term. Combining (3.20) and (3.5) we find (3.19) satisfied. We have ignored the $\nu O(h^3)$ - terms in $\ell_{\nu+1\nu}$ resp. $(\nu-1)O(h^3)$ in $\ell_{\nu\nu-1}$. That we may do so is shown by an elementary straightforward, but very lengthy computation, essentially by showing that for these $\nu O(h^3)$-terms

$$2(\nu-1)O(h^3) - \nu O(h^3) - (\nu-1)O(h^3) = O(h^3) .$$

As a consequence of (3.20),(3.21) we have proved (3.19) and therefore with (3.17), (3.18) Lemma 3.5. □

<u>Remark 3.6:</u> The conditions in (3.5) are only one possibility to enforce (3.19). The only purpose for $a' \leq 0$ and $2aq-(ap)'-\frac{a'^2}{2a} \geq 0$ is to ensure, that the h^3 [illegible] -term in (3.21) must not dominate the h^2-term in (3.21). One might choose other, less limiting, properties. We get then, with

$$q_{\nu-1} := (ap)_{\nu-1} - (ap)_0 + \int_{x_0}^{x_{\nu-1}} (\frac{a'^2}{a} - 4aq)dx ,$$

conditions of the form

$$(3.22) \quad \begin{cases} a_\nu \left(2aq - a'' - (ap)' + \frac{a'^2}{a}\right)_\nu + a'_\nu \sum_{\mu=1}^{\nu-2} (2(ap)' + \frac{a'^2}{a} - 4aq)_\mu \geq c_+, \\ \qquad\qquad \text{where } 1/\nu \leq h \quad \text{or} \\ \left(2q - \left(\frac{a'}{a}\right)' - \frac{(ap)'}{a}\right)_\nu + \frac{a'_{\nu-1}}{a^2_{\nu-1}} q_{\nu-1} \geq c_+ \quad \text{or} \\ \left(2q - \left(\frac{a'}{a}\right)' - \frac{(ap)'}{a}\right)_\nu + \frac{1}{a^2_{\nu-1}} q_{\nu-1} (a'_{\nu-1} + \frac{q_{\nu-1}}{8}) \geq c_+ . \end{cases}$$

Since the combination of these conditions with the linearized partial differential equation is even more complicated to check in concrete problems than (3.5) we confine our further discussion to (3.5). If in a special case (3.5) should not be satisfied one still can try to verify the corresponding conditions based on the inequalities given in (3.22). □

<u>Remark 3.7:</u> We have confined our discussion to $a'(\cdot) \geq 0$ in $[x_i^+,\hat{x}]$, $a'(\cdot) \leq 0$ in $[\hat{x},x_i^*]$. If the distribution of signs changes, we have to impose

$$
(3.23)\left\{\begin{array}{l}
a \in C^{3+\alpha}[x_i^+,x_i^*], p \in C^{2+\alpha}[x_i^+,x_i^*], q \in C^{1+\alpha}[x_i^+,x_i^*],\ 0 < \alpha < 1, \\
\qquad\qquad\qquad\qquad a(\cdot) \geq \underline{a} > 0 \text{ in } [x_i^+,x_i^*]. \\
\text{There exists } \hat{x} \in [x_i^+ +(k+1-s_i^+)h, x_i^*-(k+1-s_i^*)h] \text{ with} \\
a'(\hat{x}) = 0,\ a'(\cdot) \leq 0 \text{ in } [x_i^+,\hat{x}], a'(\cdot) \geq 0 \text{ in } [\hat{x},x_i^*]\ , \\
\{2aq-|(ap)'|\}(\cdot)<C_-< 0 \text{ in a neighbourhood of } \hat{x}, \\
\{2aq-a''-(\mathrm{sgn}a')(ap)' + \frac{a'^2}{a}\}(\cdot) \leq C_-< 0 \text{ in } [x_i^+,x_i^*], \\
\{2aq-(\mathrm{sgn}a')(ap)'- \frac{a'^2}{a}\}(\cdot) \leq 0 \text{ in } [x_i^+,x_i^*]\ , \\
\text{with } \mathrm{sgn}a' = \begin{cases} 1 \text{ in } (\hat{x},x_i^*]\ , \\ -1 \text{ in } [x_i^+,\hat{x}). \end{cases}
\end{array}\right.
$$

Then we find $S \leq -\delta I$, $\delta > 0$ and Lemmas 3.1-3.4 remain valid, if δ, $\geq$ is replaced by $-\delta$, $\leq$ and the main result $\| A^{-1} \| \leq \delta^{-1}$ stays unchanged. So we find again the stability result $\| A^{-1} \| \leq Ch^{-2}$, $C > 0$. □

Remark 3.8: If we want to prove the result corresponding to Lemma 2.5 for B_2 the order of the indices in (3.2)ff has to be inverted. That may be simulated by the transformation of x into -x, and in (3.20), (3.21) a' and (ap)' change into -a' and -(ap)', the rest stays unchanged. So conditions (3.5) imply Lemma 3.5 for S defined by B_1 and by B_2. □

Finally we want to point out a very important special case:

Remark 3.9: If a = const. Lemma 3.5 remains valid, if (3.5) is replaced by

$$
(3.24)\qquad \left\{\begin{array}{l}
p \in C^{2+\alpha}[x_i^+,x_i^*]\ ,\ q \in C^{1+\alpha}[x_i^+,x_i^*],\ a > 0, 0 < \alpha < 1 \\
(2q-|p'|)(\cdot) \geq c_+ > 0 \text{ in } [x_i^+,x_i^*]\ .
\end{array}\right.
$$

In (3.24) one may replace $2q-|p'|$ by one of the conditions

$$2q + p' \text{ for } x \geq \hat{x} \quad\text{and}\quad 2q-p' \text{ for } x \leq \hat{x}$$

respectively

$$2q - p' \text{ for } x \geq \hat{x} \quad\text{and}\quad 2q+p' \text{ for } x \leq \hat{x}$$

where $\hat{x}$ is an arbitrary point in $[x_i^+ +(k+1-s_i^+)h, x_i^* -(k+1-s_i^*)h]$. □

Combining Lemmas 3.1-3.5 we obtain

Theorem 3.10: Let a, p, q introduced in (2.15) satisfy one of the conditions (3.5), (3.24), (3.23) or (3.22), let $k \leq 6$ and h be small enough, and let A be the matrix defined in (2.12) - (2.15). Then there exist constants $C_k > 0$, independent of h, such that

$$\|A^{-1}\|_2 \leq C_k(\operatorname{diam}\Omega)^2 \cdot h^2. \tag{3.25}$$

If we apply Theorem 3.10 to a linear elliptic equation of the form (2.9) we obtain a stability result in the sense that the inverse operator to $\varphi_h F_0$, defined in (2.11), is uniformly bounded for $h \to 0$. For the nonlinear discretization (2.4)-(2.7) the stability is understood in the above sense for the operator $F'(y)$, where y is close to the exact solution z of (2.2) (see Stetter [26]). So this "nonlinear" stability property strongly depends on small enough $\|y - z\|$. We only formulate a stability result in this nonlinear setting for problem (2.2). We will treat the more complicated problem (2.16) with the much less restrictive conditions in §4.

Theorem 3.11: Let $k \leq 6$, h be small enough, and for small $\|y - z\|$ and in (3.26),(3.27)

$$\left\{\begin{aligned} & a_i(\cdot) \in C^{3+\alpha}(\Omega), \ p_i(\cdot) := f^{(\underbrace{0,\dots,1}_{i+2},0,\dots,0)}(\cdot,y(\cdot),\nabla y(\cdot))/a_i(\cdot) \in C^{1+\alpha}(\Omega), \\ & \qquad\qquad\qquad\qquad i = 1,2,\dots,n, \\ & q(\cdot) := f^{(0,1,0,\dots,0)}(\cdot,y(\cdot),\nabla y(\cdot))/\sum_{\nu=1}^{n} a_\nu(\cdot) \in C^{1+\alpha}(\Omega), \quad 0 < \alpha < 1. \end{aligned}\right. \tag{3.26}$$

Further, let for each maximal intersecting interval $[x_i^+, x_i^*] \subset \bar{\Omega} \cap G_{h,n}$ (that is $x_i^+, x_i^* \in \partial\Omega$) in the direction of the co-ordinate vector e_i exist $\hat{x}_i \in [x_i^+ + (k+1-s_i^+)h, \ x_i^* - (k+1-s_i^*)h]$ such that

$$\left\{\begin{aligned} & a_i'(\hat{x}_i) = 0, \ a_i'(\cdot) \geq 0 \ \text{in} \ [x_i^+, \hat{x}_i], \ a_i'(\cdot) \leq 0 \ \text{in} \ [\hat{x}_i, x_i^*], \\ & \{2a_i q - |(a_i p_i)'|\}(\cdot) \geq c_+ > 0 \ \text{in an interval of length} \geq h \ \text{with} \\ & \qquad\qquad\qquad\qquad \text{middlepoint } \hat{x}, \\ & \{2a_i q - a_i'' + (\operatorname{sgn} a_i')(a_i p_i)' + \frac{a_i'^2}{a_i}\}(\cdot) \geq c_+ > 0 \ \text{in} \ [x_i^+, x_i^*] \\ & \{2a_i q + (\operatorname{sgn} a_i')(a_i p_i)' - \frac{a_i'^2}{2a_i}\}(\cdot) \geq 0 \ \text{in} \ [x_i^+, x_i^*]. \end{aligned}\right. \tag{3.27}$$

Then the discretization, defined in (2.4)-(2.7), is stable in the sense discussed above. Instead of (3.27) one could define the corresponding conditions based on [illegible]) or more complicated conditions based on (3.[illegible]).

We want to formulate the corresponding result for constant a_i.

Theorem 3.12: Let $k \le \delta$, h be small enough, a_i = const and for small enough $\|y - z\|$ and smooth enough y

$$(3.28)\qquad \begin{cases} p_i(\cdot) := f^{(\underbrace{0,\dots,1}_{i+2},\dots 0)}(\cdot, y(\cdot), \nabla y(\cdot)) \in C^{2+\alpha}(\bar\Omega), \\ q(\cdot) := f^{(0,1,0,\dots,0)}(\cdot, y(\cdot), \nabla y(\cdot)) \in C^{1+\alpha}(\bar\Omega), \qquad 0 < \alpha < 1, \\ 2q(\cdot) - \sum_{i=1}^{n} |p_i'(\cdot)| \ge c_* > 0 \quad \text{in } \Omega. \end{cases}$$

Then the discretization, defined in (2.4)-(2.7) is stable.

Proof: Theorem 3.11 is an immediate consequence of Theorem 3.10. Theorem 3.12 is derived from condition (3.24). By

$$q_i^*(\cdot) := p_i'(\cdot)/2 + c_*/n, \qquad i = 1,\dots,n,$$

$$r(\cdot) := q(\cdot) - \sum_{i=1}^{n} q_i^*(\cdot) \ge 0,$$

$$q_1 := q_1^* + r, \quad q_i := q_i^*, \qquad i = 2,\dots,n$$

we have found p_i, q_i, $i = 1,\dots,n$, satisfying (3.24). □

For the special case of the Poisson equation

$$(3.29)\qquad \begin{cases} p_i(\cdot) = 0, \quad i = 1,\dots,n, \quad q(\cdot) = 0, \quad \text{that is} \\ F_p y := \begin{cases} \Delta y + f(\cdot) = -\sum_{i=1}^{n} y_{x_i x_i}(\cdot) + f(\cdot) \quad \text{in } \Omega \\ y(\cdot) - q(\cdot) \text{ on } \partial\Omega, \end{cases} \end{cases}$$

the stability result corresponding to Theorem 3.2 has been proved in Pereyra-Proskurowski-Widlund [24]. We can not include this special case in our general Theorem 3.8, since (3.27) would correspond to $c_* = 0$ in (3.24). The proof given above breaks down for $c_* = 0$.

4. Stability for $k \le 4$ under less stringent conditions.

For $k \le 4$ a much easier stability proof may be given. If P is a $\mu \times \mu$ permutation matrix, then $P^T B P$ and B contain the same diagonal elements only in a changed order. Further, the elements found in the same row and column with b_{ii} will be found in that row and column which are [illegible] by the same [illegible]ition $\bar{i}$ [illegible] the element [illegible] after the permutation. With this [illegible] the [illegible] Lemma [illegible] obvious:

Lemma 4.1: Let P be a permutation matrix. Then $B = (b_{ij})$ is diagonal dominant if and only if $P^T B P = (\overline{b_{ij}})$ is diagonal dominant and

$$\min_{i=1}^{\mu} \{b_{ii} - \sum_{\substack{j=1\\ j\neq i}}^{\mu} |b_{ij}|\} = \min_{i=1}^{\mu} \{\overline{b_{ii}} - \sum_{\substack{j=1\\ j\neq i}}^{\mu} |\overline{b_{ij}}|\} .$$

Further we need the elementary

Lemma 4.2: Let $A = \sum_{j=1}^{\rho} B_j = (a_{i\ell})$, let each nontrivial row in $B_j = (b_{ji\ell})$ be diagonal dominant with positive diagonal elements. Further let for each row of A exist a nontrivial row with the same index in at least one of the B_j. Then A is diagonal dominant and

$$\min_{i=1}^{\mu} \{a_{ii} - \sum_{\substack{\ell=1\\ \ell\neq i}}^{\mu} |a_{i\ell}|\} \ge \min_{j=1}^{\rho} \left\{ \min_{\substack{i=1\\ b_{jii}\neq 0}}^{\mu} \{ b_{jii} - \sum_{\substack{\ell=1\\ \ell\neq i}}^{\mu} |b_{ji\ell}|\}\right\} .$$

Now if we want to prove stability it is enough to show that the matrices defined in (2.14) are diagonal dominant. For that purpose we need less stringent conditions as in §3 (see (3.5)).

Lemma 4.3: Let in (2.9) a_i, b_i, c satisfy

$$(4.1)\quad \begin{cases} a_i, b_i, c \in C(\Omega), \quad 0 < \underline{a} \le a_i(\cdot) \\ 0 < Q_* \le q(\cdot) := c(\cdot) / \sum_{i=1}^{n} a_i(\cdot) \le Q^* \\ \left\| \dfrac{b_i(\cdot)}{a_i(\cdot)} \right\|_{\infty} \le P_i^* , \quad i = 1,\dots,n . \end{cases}$$

Then for $h < \min_{i=1}^{n} \{2/P_i^*\}$ and $k = 1,2,3,4$ the matrix $B = (b_{ij})$ in (2.14) is diagonal dominant with

$$(4.2)\qquad \min_{i=0}^{\ell} \{b_{ii} - \sum_{\substack{j=0\\ j\neq i}}^{\ell} |b_{ij}|\} \ge c_+ h^2 , \quad c_+ > 0 .$$

Proof: Under the assumption (4.1) and for $h < \min_{i=1}^{n} (2/P_i^*)$ we have with $q :=$ $c(\cdot)/\sum_{i=1}^{n} a_i(\cdot)$ and $p(\cdot) := b_i(\cdot)/a_i(\cdot)$ (see (2.15))

$$2 + h^2 q_\nu - (|-1 - \frac{h}{2} p_\nu| + |-1 + \frac{h}{2} p_\nu|) = h^2 q_\nu \geq h^2 Q_* .$$

Further, the $\frac{\alpha_\nu'}{\alpha_0'}$ and $\frac{\alpha_\nu}{\alpha_0}$, $\nu = 1,\ldots,k$, have alternating sign for $0 \leq s \leq 1$ and $\frac{\alpha_1}{\alpha_0} > 0$. So we find that, under the assumption (4.1) and for $0 \leq s \leq 1$,

$$1 + h^2 q + \frac{\alpha_1'}{\alpha_0'}(1 + \frac{h}{2} p_0) - (|-1 + \frac{h}{2} p_0 + \frac{\alpha_2'}{\alpha_0'}(1 + \frac{h}{2} p_0)| + \ldots + |\frac{\alpha_k'}{\alpha_0'}(1 + \frac{h}{2} p_0)|)$$

$$= 1 + h^2 q + \frac{h}{2} p_0 + (1 + \frac{h}{2} p_0)\left(\frac{\alpha_1' + \alpha_2' - \alpha_3' + \alpha_4' - + \ldots + (-1)^k \alpha_k'}{\alpha_0'}\right) \geq c_* h^2$$

for all $h > 0$

if and only if,

$$g_k(s) := 1 + \frac{\alpha_1' + \alpha_2' - \alpha_3' + \ldots + (-1)^k \alpha_k}{\alpha_0'} \geq 0 \quad \text{for } 0 \leq s \leq 1 .$$

Now one verifies either by straightforward discussion or by using computers that

$$\min_{0 \leq s < 1} g_k(s) \begin{cases} > 0 & \text{for } k = 1,2,3,4. \\ < 0 & \text{for } k = 5,6. \end{cases}$$

So (4.2) holds just for $k = 1,2,3,4$ and the Lemma is proved.

Combining Lemmas 4.1 to 4.3 we obtain by Gerschgorin's Theorem [21]:

Theorem 4.4: Let (4.1) be satisfied and $h < \min_{i=1}^{n} \{2/P_i^*\}$. Then the matrix A is regular and there are positive constants D_k such that

$$\|A^{-1}\|_\infty \leq D_k \cdot h^{-2} .$$

In contrast to Theorem 3.7 we have an estimation for $\|A^{-1}\|_\infty$ instead of $\|A^{-1}\|$.

Since in the proofs of our Lemmas 4.1-4.3 the coefficients a_i in (2.1) only had to satisfy (4.1) and since we obtain stability for the nonlinear problem simply by proving stability for the linear problem we can generalize our Theorem 4.4 immediately to the more general case given in (2.16), (2.17):

__Theorem 4.5__: __Let__ z __be the unique solution of__ (2.16), $h < \min_{i=1}^{n} \{2/P^*\}$, __and let__

$$(4.3)\quad \begin{cases} 0 < \underline{a} \leq a_i(\cdot,z(\cdot),z_{x_i}(\cdot)) \ \underline{\text{in}}\ \Omega\ , \quad i = 1,2,\dots,n\ , \\ 0 < Q_* \leq \{f^{(0,1,0\dots0)}(\cdot,z(\cdot),\nabla z(\cdot)) - \sum_{i=1}^{n} a_i^{(0,1,0)}(\cdot,z(\cdot),z_{x_i}(\cdot))z_{x_ix_i}(\cdot)\} \\ \qquad \times(\sum_{i=1}^{n} a_i(\cdot,z(\cdot),z_{x_i}(\cdot)))^{-1} \leq Q^* \ \underline{\text{in}}\ \Omega\ , \\ \underline{\text{and}}\ ,\ \underline{\text{with}}\ P_i^*\ \underline{\text{in}}\ (4.1),\ \underline{\text{let}} \\ |\{f^{\underbrace{(0,\dots,1}_{i+2},0\dots0)}(\cdot,z(\cdot),\nabla z(\cdot)) - a_i^{(0,0,1)}(\cdot,z(\cdot),z_{x_i}(\cdot))z_{x_ix_i}(\cdot)\} \\ \qquad\qquad \times(a_i(\cdot,z(\cdot),z_{x_i}(\cdot)))^{-1}| \leq P_i^* \ \underline{\text{in}}\ \Omega\ , \\ \underline{\text{with continuous functions}}\ \ a_i,\ f^{(0,1,\dots,0)}(\cdot,z(\cdot),\nabla z(\cdot))\ \ \underline{\text{and}} \\ f^{\underbrace{(0,\dots,01}_{i+2},0,\dots,0)}(\cdot,z(\cdot),\nabla z(\cdot)) \quad \underline{\text{in}}\ \Omega\ . \end{cases}$$

__Then the discretization of__ F __in__ (2.16) __is stable in a neighbourhood of the exact solution__ z.

5. Convergence and asymptotic expansion.

With the stability, proved in §3 and 4, we obtain convergence and asymptotic expansions by studying the local discretization error. Before we do so we need some formal notations. With the spaces $E := C^2(\Omega) \times C(\bar{\Omega})$ and $E^0 := C(\Omega) \times C(\partial\Omega)$ and the grid and [illegible], we define

$$(5.1)\quad \begin{cases} E_h := E_h^0 := \{\eta_h : D_{h,\Omega} := (\Omega_{h,n} \cap \Omega) \cup (G_{h,n} \cap \partial\Omega) \to \mathbb{R}\} \\ \text{with one of the norms} \\ \qquad \|\eta_h\|_{h,\infty} := \|\eta_h\|_{h,\infty}^0 := \max\limits_{x \in D_{h,\Omega}} |\eta_h(x)| \qquad \text{or} \\ \qquad \|\eta_h\|_{h,2} := \|\eta_h\|_{h,2}^0 := \Big(h^n \sum\limits_{x \in D_{h,\Omega}} |\eta_h(x)|^2 \Big)^{1/2} \\ \qquad \text{and the restriction operators} \\ \qquad \Lambda_h := \begin{cases} E \to E_h \\ y \mapsto y|_{D_{h,\Omega}} \end{cases}, \quad \Lambda_h^0 : \begin{cases} E^0 \to E_h^0 \\ (u,v) \mapsto (u|_{\Omega_{h,n} \cap \Omega}, v|_{G_{h,n} \cap \partial\Omega}). \end{cases} \end{cases}$$

Now we have (for B see (2.1))

Lemma 5.1: Let in (2.2) the solution $z \in C^{2(q+1)+\alpha}(\bar{\Omega})$, $\frac{\partial^\nu}{\partial z_1^{\nu_1} \dots \partial z_n^{\nu_n}} f := f^{(0,0,\nu_1,\dots,\nu_n)}$ and $\frac{\partial^\nu}{\partial z_1^{\nu_1} \dots\dots \partial z_n^{\nu_n}} f(\cdot,\cdot,\cdot) \in C^\alpha(\bar{\Omega} \times B)$ for $\nu_1 + \dots + \nu_n = \nu = 1,2,\dots,q$, and let the usual formal differential operator in the multivariate Taylor expansion be given as $\left(\sum\limits_{i=1}^{n} \delta_i \frac{\partial}{\partial z_i}\right)^\nu$. Then the local discretization error $(\varphi_h F)(\Lambda_h z)$ is given in regular points x as

$$(5.2)\quad \begin{cases} h^2(\varphi_h F)(\Lambda_h z)(x) = \sum\limits_{i=1}^{n} \Big\{- a_i(x) \sum\limits_{j=2}^{q+1} \frac{2h^{2j}}{(2j)!} \frac{\partial^{2j} z(x)}{\partial x_i^{2j}} \Big\} \\ +h^2 \sum\limits_{\nu=1}^{q} \frac{1}{\nu!} \Big\{ \sum\limits_{i=1}^{n} \Big(\sum\limits_{j=1}^{q} \frac{h^{2j}}{(2j+1)!} \frac{\partial^{2j+1} z(x)}{\partial x_i^{2j+1}} \Big) \frac{\partial}{\partial z_i} \Big\}^\nu f(x,z(x),\nabla z(x)) \\ + O(h^{2(q+1)+\alpha}) \text{ for } x \in \Omega_h. \end{cases}$$

In irregular mesh points $\sum\limits_{j=2}^{q+1}$ and $\sum\limits_{j=1}^{q}$, summing the derivatives of z, have to be changed in the following way:

Let $z^* \in C^{2(q+1)+\alpha}(\Omega^*)$ be a smooth extension of z, where $\Omega^* \supset \Omega$ is large enough to contain all grid neighbours for irregular points. Then use

$$(5.3)\quad \begin{cases} \sum\limits_{j=2}^{q+1} \frac{2h^{2j}}{(2j)!} \frac{\partial^{2j} z(x)}{\partial x_i^{2j}} + \frac{s_i}{k+1} h^{k+1} \frac{\partial^{k+1} z^*(\xi_i)}{\partial x_i^{k+1}} \ \underline{\text{resp.}} \\ \sum\limits_{j=1}^{q} \frac{h^{2j+2}}{(2j+1)!} \frac{\partial^{2j+1} z(x)}{\partial x_i^{2j+1}} + \frac{s_i}{2(k+1)} h^{k+2} \frac{\partial^{k+1} z^*(\bar{\xi}_i)}{\partial x_i^{k+1}} \\ \underline{\text{with}}\ \xi_i, \bar{\xi}_i\ \underline{\text{on the gridline through the point}}\ x \\ \underline{\text{in}}\ e_i\ \underline{\text{direction and}}\ i\text{-th}\ \underline{\text{co-ordinate in}} \\ \qquad (x \pm (k-1)he_i,\ x \mp h e_i) \end{cases}$$

<u>for the corresponding</u> $\sum\limits_{j=2}^{q+1}$ <u>resp.</u> $\sum\limits_{j=1}^{q}$ <u>in</u> (5.2).

<u>**Proof**</u>: Immediately we find (5.4) with

$$(5.4)\quad \begin{cases} \frac{u(x+h) - u(x-h)}{2h} = u'(x) + \sum\limits_{j=1}^{q-1} \frac{h^{2j}}{(2j+1)!} u^{(2j+1)}(x) + O(h^{2q-1+\alpha}) \\ \text{and} \\ \frac{u(x+h) - 2u(x) + u(x-h)}{h^2} = u''(x) + \sum\limits_{j=2}^{q} \frac{2h^{2(i-1)}}{(2j)!} u^{(2j)}(x) + O(h^{2q-2+\alpha}) \\ \text{for any } u \in C^{2q+\alpha}(x-h,x+h). \end{cases}$$

Now $z \in C^{2(q+1)+\alpha}(\bar{\Omega})$ may be extended to a $z^* \in C^{2(q+1)+\alpha}(\Omega^*)$. Such extensions exist , but they usually do not satisfy an extended differential equation (2.2). Further, since we can split (5.2) into terms, each one only involving derivatives with respect to one **variable, we may** confine the discussion now to one variable. Since we need the local discretization error, we have to replace y_1 in (2.5) by z_1, obtained as $z_1 = P_k(x+he_i)$, where P_k is defined by $P_k(x-(\nu-1)he_i) = z_{1-\nu}$, $\nu = 1,\ldots,k$, $P_k(x_i^*) = g(x_i^*)$. So we have, by the well known error formulas for polynomial interpolation,

$$(5.5)\quad \begin{cases} z(x+he_i) - z_1 = \frac{(x+he_i - x_i^*)(x+he_i - x)\ldots(x+he_i - x + (k-1)he_i)}{(k+1)!} \cdot z^{(k+1)}(\xi) \\ \qquad = \frac{s_i}{k+1} h^{k+1} \frac{\partial^{k+1} z(\xi_i)}{\partial x_i^{k+1}} . \end{cases}$$

By inserting (5.5) into (5.4) we evidently obtain (5.2), (5.3). □

Combining formulas (5.2) and (5.3) and observing that there are only $O(h^{-(n-1)})$ irregular points one obtains

Lemma 5.2: The local discretization error $(\varphi_h F)(\Delta_h z)$ admits an asymptotic expansion of the form

$$(5.6)\quad \begin{cases} \| (\varphi_h F)(\Delta_h z) - \Delta_h^o \sum\limits_{\iota=1}^{\bar{q}} h^{2\iota} g_{2\iota} \|_{h,\sigma} = O(h^{k-1+1/\sigma}) \\ \text{for } 2\bar{q} < k-1+1/\sigma, \text{ and } \sigma=2,\ k \le 6 \text{ or } \sigma = \infty,\ k \le 4, \end{cases}$$

where the $g_{2\iota}$ are defined in (5.2), (5.3).and $\bar{q} \le q$ as in Lemma 5.1.

Again, since (5.1) can be split into the terms mentioned in (2.11), we may use the arguments in [7] to prove Theorem 5.3. In this Theorem we only give conditions for a_i, f and $g, \partial\Omega$ to ensure $z \in C^{2q+2+\alpha}(\bar{\Omega})$. The more or less "geometric" type conditions, given e.g., in Agmon-Douglis-Nirenberg [1], §7 are only referred to. If they are valid, the smoothness results for the $e_{2\iota}$ are ensured.

Theorem 5.3: Let, for $k \le 4$ and $k \le 6$, the $\|\cdot\|_\sigma$-norm, $\sigma = \infty$ and $\sigma=2$, the condition (4.3) and (3.26)-(3.27) or (3.28) be valid respectively.Let further $z \in C^1(\bar{\Omega}), a_i \in C^{2q+\alpha}(\bar{\Omega})$, $f(\cdot,\cdot,\cdot) \in C^{2q+\alpha}(\bar{\Omega}\times B)$, $\partial\Omega \in C^{2q+2+\alpha}$, $g \in C^{2q+2+\alpha}(\partial\Omega)$ for $0 < \alpha < 1$ and $k - 1 + 1/\sigma \le 2q+\alpha$, and "geometric" conditions (see [1], §7) be satisfied.

Then $z \in C^{2q+2+\alpha}(\Omega)$ and we have for the approximate solution ζ_h

$$(5.7)\quad \begin{cases} \zeta_h - \Delta_h z = \Delta_h \sum\limits_{\iota=1}^{\tilde{q}} h^{2\iota} e_{2\iota} + O(h^{k-1+1/\sigma}) \\ \text{for } \tilde{q} = \max\{m \in \mathbb{N} \mid 2m < k-1+1/\sigma,\ m \le q\}\ . \end{cases}$$

Here $O(h^{k-1+1/\sigma})$ refers to the difference with respect to $\|\cdot\|_\sigma, \sigma=2,\infty$.Further we have

$$e_{2\iota}(\cdot) \in C^{2(q-\iota)+2+\alpha}(\bar{\Omega}),\ e_{2\iota}(\cdot) = 0 \text{ on } \partial\Omega\ ,\quad \iota = 1,2,\dots,\tilde{q}\ .$$

(5.7) corresponds for (3.29) to the result of Pereyra-Proskurowski-Widlund [24].

A result,given in Agmon-Douglis-Nirenberg [1], §7, may sometimes help to avoid difficulties arising from the violation of $\partial\Omega \in C^{2q+2+\alpha}$ in isolated points. Since in all our proofs we only need differentiability properties along grid lines, one only has to choose the grid such that its lines avoid non-smooth boundary points. Then the results in [1], §7 may be used to obtain results for the case $h_0 \ge h \ge h_1 > 0$, which might sometimes be useful.

6. Two special cases.

The breakdown of the asymptotic expansion in Theorem 5.4 is, for a sufficiently smooth situation, caused by the interpolation error, given in (5.5). This error will be less disturbing in one of the two following cases:

(6.1) Let $s_i = 0$, $i = \mu_1,\dots,\mu_m$ (see (2.6)), in (5.5) for all $x \in \Omega_{h,i}$.

(6.2) We admit only h, such that

s_i = const. (independent of h), $i = \mu_1,\dots,\mu_m$, in (5.5) for all $x \in \Omega_{h,i}$.

The second case is, admittedly, somewhat artificial, but if it is satisfied we have instead of (5.7) the relation

$$(6.3)\quad \begin{cases} \zeta_h - \Delta_h z = \Delta_h \sum\limits_{\iota=1}^{q^+} h^{2\iota} e_{2\iota} + O(h^{k+1/\sigma}) \\ \text{if (6.2) is satisfied, with } k+1/\sigma \le 2q+\alpha \text{ and} \\ q^+ = \max\{m \in \mathbb{N} \mid 2m < k+1/\sigma,\ m \le q\}. \end{cases}$$

In some cases, when (6.2) is satisfied we may, by a proper choice of (different) stepsizes in the directions of different co-ordinate axises,even obtain $s_i = 0$, that is (6.1). This case has been treated in Pereyra [23] for uniformly elliptic operators of the following (casilinear) type

$$(6.4)\quad F_c y: \left\{ \begin{array}{ll} -\sum\limits_{i,j=1}^{n} a_{ij}(\cdot) y_{x_i x_j}(\cdot) + f(\cdot, y(\cdot), \nabla y(\cdot)) & \text{in } \Omega \\ y(\cdot) - g(\cdot) \text{ on } & \partial\Omega \end{array} \right\}$$

with $\sum\limits_{i,j=1}^{n} a_{ij}(\cdot)\xi_i\xi_j \ge \alpha \|\xi\|_2^2$ in Ω with $\alpha > 0$ and for any real vector $\xi^T = (\xi_1,\dots,\xi_n)$. The discretization of (6.4) under the condition (6.1) is obtained by using (2.4) in regular and irregular mesh points and by replacing in irregular mesh points the $\eta_h(x \pm he_i)$ with $x \pm h e_i \in \partial\Omega$ by the boundary values $g(x \pm he_i)$. Then Pereyra [23] has proved (the smoothness properties of the $e_{2\iota}$ in (6.5) may be checked by some additional computations).

Theorem 6.1: *Let* (6.1) *be satisfied and the operator* F_c *in* (6.4) *be discretized as described above. Further let* $a_{ij} \in C^{2q+\alpha}(\bar\Omega)$, $f(\cdot,\cdot,\cdot) \in C^{2q+\alpha}(\bar\Omega \times B)$ *and* $\partial\Omega \in C^{2q+2+\alpha}$ $g \in C^{2q+2+\alpha}(\partial\Omega)$, $z \in C^1(\bar\Omega)$ *and "geometric" conditions for* $\partial\Omega$ *be satisfied.Then* $z \in C^{2q+2\alpha}(\bar\Omega)$ *and*

$$(6.5)\quad \zeta_h - \Delta_n z = \Delta_h \sum_{\iota=1}^{q} h^{2\iota} e_{2\iota} + O(h^{2q+\alpha}),\ e_{2\iota}(\cdot) \in C^{2(q-\iota)+2+\alpha}(\bar\Omega).$$

7. Discrete Newton methods and deferred corrections

As an immediate consequence of the results in §§5,6 we may use standard extrapolation techniques to get higher order results. The general approach is given in Bulirsch-Stoer [13] and applied to the Dirichlet problem for the Poisson equation in Pereyra-Proskurowski-Widlund [24]. So we only give the result here.

Theorem 7.1: By applying Richardson extrapolation once resp. twice to the methods given in Theorem 5.3 for $k = 4$ resp. $k = 6$ we get approximations of order 3.5 resp. 5.5 in ℓ_2-norm. If (6.2) is valid orders 4.5 resp. 6.5 are obtainable by two resp. [illegible] three extrapolation steps. Even higher orders (up to $2q+1$, see Theorem 6.1) are obtainable, if (6.1) is satisfied.

Using Richardson extrapolation requires the solution of the nonlinear system $(\varphi_h F)\zeta_h = 0$ for decreasing h. That implies, for higher dimensions n, strongly increasing the sizes of the systems of nonlinear equations to be solved. We are able to avoid this disadvantage by either applying Pereyra's [22] iterated deferred corrections or iterated defect corrections or descrete Newton methods [6,8].

The first two methods require for every correction step the solution of a nonlinear equation (of the same size) where the original right hand side 0 has to be changed. In discrete Newton methods only the original nonlinear equation has to be solved. The corrections are obtained by solving linear systems (of the same size)with a fixed matrix and different right hand sides and therefore need much less computational time.

Most of the conditions for the applicability of iterated defect corrections are easily verified, whereas for discrete Newton methods we need a lot of technical formalism. So we prove the results for iterated defect corrections and present discrete Newton methods and sketch deferred corrections without proof. For deferred corrections we have to approximate (only k=6 is discussed here) the local discretization error $(\varphi_h F)\Delta_h z$ given in (5.2) as $(\varphi_h F)(\Delta_h z)(x)=h^2 b_1(x) + h^4 b_2(x) + \dots$.

Now we have from (5.2), with $\frac{\partial^\nu}{\partial z_1^{\nu_1}\dots\partial z_n^{\nu_n}} f := f^{(0,0,\nu_1\dots,\nu_n)}, \nu_1+\dots+\nu_n=\nu$,

$$h^2 b_1(x) = \frac{h^2}{6} \sum_{i=1}^{n} \{-\frac{a_i(x)}{2} \frac{\partial^4 z(x)}{\partial x_i^4} + \frac{\partial^3 z(x)}{\partial x_i^3} \frac{\partial}{\partial z_i} f(x,z(x),\nabla z(x))\}$$

and

$$h^4 b_2(x) = \frac{h^4}{120} \Big[\sum_{i=1}^{n} \{- \frac{a_i(x)}{3} \frac{\partial^6 z(x)}{\partial x_i{}^6} + \frac{\partial^5 z(x)}{\partial x_i{}^5} \frac{\partial}{\partial z_i} f(x,z(x),\nabla z(x))\}$$

$$+ \frac{5}{3} (\sum_{i=1}^{n} \frac{\partial^3 z(x)}{\partial x_i{}^3} \frac{\partial}{\partial z_i})^2 f(x,z(x),\nabla z(x))\Big]$$

and so we need approximations for $h^2 b_1(x)$ and $h^2 b_1(x) + h^4 b_2(x)$ of the order 4 and 6 respectively. Now the same remark ist appropriate for irregular points with respect to $s \to 1$ for the behaviour of the α_ν/α_o as given in Pereyra-Proskurowski-Widlund [24], p. 9.

In iterated defect corrections and in discrete Newton methods we have to discretize equations of the form

(7.1) $\quad F_d y: =F_o y+d$, with $d: =\begin{pmatrix} e \text{ in } \Omega \\ e^* \text{on } \partial\Omega \end{pmatrix}$.

For defect corrections F_o is given as F in (2.1), whereas $F_o \in L(C^2(\Omega) \cap C(\bar{\Omega}), C(\Omega)\times C(\partial\Omega))$ for discrete Newton methods. With the discretization, defined in 2., we obtain

$$(7.2) \quad \begin{cases} (\varphi_h F_d)\ \eta_h =(\varphi_h F_o)\eta_h + \Omega_h d, \\ \text{where } \varphi_h F_o \text{ is described in §2 and} \\ \Omega_h \in L(C(\Omega)\times C(\partial\Omega),(\Gamma_{h,n}\cap\Omega)\times(G_{h,n}\cap\partial\Omega)). \end{cases}$$

For the considerations below we have

(7.3) $\quad d = \begin{pmatrix} e \text{ in } \Omega \\ 0 \text{ on } \partial\Omega \end{pmatrix}$ and therefore $\Omega_h d = \begin{pmatrix} e|_{\Gamma_{h,n}\cap\Omega} \\ 0 \end{pmatrix}$,

whereas for $e^* \neq 0$ in (7.1) the component of $\Omega_h d$ in $\Gamma_{h,n}\cap\Omega$ is more complicated since the boundary values enter the difference equations for irregular points in (2.6). However, in either case $\Omega_h d$ is independend of η_h. Discretizations, for which (7.1) implies (7.2), are called strongly linear, see Böhmer [8]. In [8] we have proved that , if $Fy + d = 0$ is uniquely solvable for small enough $\| d \|$ and if the strongly additive discretization method generates $\Omega_h d$ independend of η_h, a stable $\varphi_h F$ and an error asymptotic, where the coefficients $e_{2\iota}$ in (5.7) are elements of linear spaces with certain properties, see

Assumption 2.4 in [8] part II, the iterated defect correction provides us with the results given below. All conditions mentioned above, except Assumption 2.4, have been shown to be satisfied. Assumption 2.4 is, for our elliptic problem, a consequence of the smoothness requirements in Theorem 5.3, which we do not want to formulate explicetly and of the structure of F in (2.1), see [8], part I.

For discrete Newton methods and iterated defect corrections we have to compute defects in the form

(7.4) $$d = \pm Fz_{\ell-1},\ \ell = 1,2,\dots .$$

That means, we need extension operators which extend discrete approximations, as η_h and ζ_h, defined on $(\Gamma_{h,n}\cap\Omega)\times(G_{h,n}\cap\partial\Omega)$, to continuous functions $y_h \in E$, such that Fy_h or at least $\overset{\circ}{\Delta}_h Fy_h$ is defined. This may be done in different ways, all based on the observation that it is enough to use univariate interpolation or approximation operators T. For a point $x \in \Omega_h$ we then find along a grid line in i-th coordinate direction the situations in Figure 3:

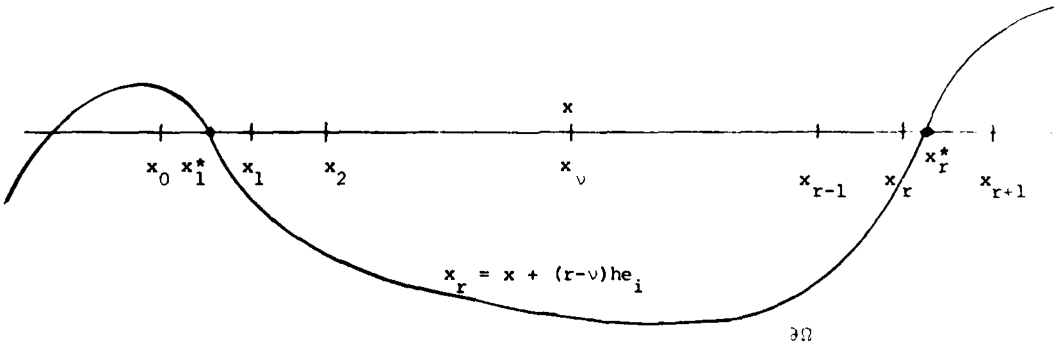


Figure 3

In $x_1,\dots,x_r$ we know the corresponding $\eta_1,\dots,\eta_r$ (as the approximate values given by the difference equations (2.4)-(2.8)). Here r depends on x and the direction e_i. Further we know in x_1^* and x_r^* the η_1^* and η_r^* from the boundary condition. Finally we may compute η_0 and η_{r+1} by the appropriate formulas (2.5). Let

$$\eta_h^T := (\eta_0,\eta_1,\dots,\eta_r,\eta_{r+1}) .$$

Method 1: Continuous piecewise polynomials:

With a $\lambda \geq k$, k as in (2.5), define T as (Π_λ are the polynomials of degree λ and [s] = entire part of s)

$$(7.5)\quad \begin{cases} P_\lambda := T\eta_h \big|_{(x_{\mu\lambda}, x_{(\mu+1)\lambda})} \in \Pi_\lambda, \mu=0,1,\ldots,\mu^* := [\frac{r+1}{\lambda}] - 1, \\ P_\mu(x_i) = \eta_i,\ i = \mu\lambda,\ldots,(\mu+1)\lambda, \\ P_{\mu^*+1} := T\eta_h \big|_{(x_{(\mu^*+1)\lambda}, x_{r+1})} \\ P_{\mu^*+1}(x_i) = \eta_i,\ i = r+1-\lambda,\ldots,r+1. \ \square \end{cases}$$

Method 2: Spline-functions:

Define T as a spline-function of degree $\lambda \geq k$ as discussed in [4,6,8] . □

Method 3: Overlapping polynomials (discrete splines):

Let $\lambda = \lambda^* + 2\kappa \geq k$, $\kappa = 1,2,3,4$ and define T as

$$(7.6)\quad \begin{cases} P_o := T\eta_h \big|_{(x_o, x_{\lambda^*+\kappa})} \in \Pi_\lambda, \\ P_o(x_\nu) = \eta_\nu, \nu = 0,\ldots,\lambda, \\ P_\mu := T\eta_h \big|_{(x_{\mu\lambda^*+\kappa}, x_{(\mu+1)\lambda^*+\kappa})} \in \Pi_\lambda, \mu = 1,\ldots,[\frac{r+\kappa}{\lambda^*}] - 1, \\ P_\mu(x_\nu) = \eta_\nu,\ \nu = \mu\lambda^*,\ldots,(\mu+1)\lambda^*+2\kappa,\ \mu = 1,\ldots,[\frac{r+\kappa}{\lambda^*}] - 1. \\ \text{The last polynomial is defined analogously to (7.5)} \ \square \end{cases}$$

Method 4: Symmetric formulas and polynomial extrapolation [8,24] :

Interpolate the points (x_ν, ξ_ν), for $\nu = 0,1,\ldots,k$, $\nu = r-k+1,\ldots,r+1$, $\nu = 1,\ldots,\lambda+1$, and $\nu = r-\lambda,\ldots,r$ by polynomials $P_{k,i}$, $P_{k,e}$, $P_{\lambda,i}$ and $P_{\lambda,e}$, respectively, and compute extrapolated provisional values

$$(7.7)\qquad \eta_\nu = P_{k,i}(x_\nu), \nu = -1,-2,\ldots \text{ and } \eta_\nu = P_{k,e}(x_\nu),\ \nu = r+2, r+3,\ldots,$$

and

$$(7.8)\qquad \eta_\nu = P_{\lambda,i}(x_\nu), \nu = 0,-1,-2,\ldots \text{ and } \eta_\nu = P_{\lambda,e}(x_\nu), \nu = r+1, r+2,\ldots,$$

respectively. Then use for $\ell = 1,2,3$ symmetric formulas, based on 5,7,9 points to compute the defect $Fz_{\ell-1}$ in (7.4). For points close to the boundary, outer points are necessary to compute these approximations. The outer values are obtained as provisional values from polynomial extrapolation and have to be updated after every iteration as far as necessary. □

For other possibilities see [6].

To formulate iterated defect corrections we have to replace F in (2.1) by $F_{\ell-1}$ defined as

$$F_{\ell-1}y:=Fy-Fz_{\ell-1},\ell=1,2,\dots .$$

In the discretization we need $Fz_{\ell-1}$ only in $(\Gamma_{h,n}\cap\Omega)\cup(G_{h,n}\cap\partial\Omega)$. Since the $z_{\ell-1}$ satisfy the boundary condition in the grid points $G_{h,n}\cap\partial\Omega$, up to round-off-errors, we are in the situation (7.3) and obtain

$$(7.9)\quad\begin{cases}(\varphi_hF_{\ell-1})\eta_h(x)=(\varphi_hF)\eta_h(x)-Fz_{\ell-1}(x)\text{ for }x\in\Omega_h\cup\Omega_{h,i}.\\ \\ (\varphi_hF_{\ell-1})\eta_h(x)=\eta_h(x)\text{ for }x\in G_{h,n}\cap\partial\Omega.\end{cases}$$

Now the iterated defect corrections are defined as follows

$$(7.10)\quad\begin{cases}\text{(i)}\quad\text{given the starting value }\zeta_{h,o}=\zeta_h\text{ such that }(\varphi_hF)\zeta_{h,o}=0\\ \qquad\text{compute }z_{\ell-1}=T\zeta_{h,\ell-1}\text{ and }\xi_{h,\ell-1}\text{ from}\\ \text{(ii)}\quad(\varphi_hF_{\ell-1})\xi_{h,\ell-1}=0\text{ with }\varphi_hF_{\ell-1}\text{ in (7.9)},\\ \text{(iii)}\quad\text{finally, define }\zeta_{h,\ell}:=\zeta_{h,o}-(\xi_{h,\ell-1}-\zeta_{h,\ell-1}),\ell=1,2,\dots .\end{cases}$$

In discrete Newton methods we have to discretize linear equations of the form

$$(7.11)\qquad F'(z_o)(z_\ell-z_{\ell-1})=-Fz_{\ell-1}.$$

So we need, see (2.4) and (2.10),

$$(7.12)\quad\begin{cases}(\varphi_hF'(z_0))\tau_h(x)=-\sum\limits_{i=1}^{n}a_i(x)\dfrac{\tau_h(x+he_i)-2\tau_h(x)+\tau_h(x-he_i)}{h^2}\\ +\sum\limits_{i=1}^{n}f^{(\underbrace{0,\dots,1}_{i+2},0\dots0)}(x,z_0(x),\nabla z_0(x))\dfrac{\tau_h(x+he_i)-\tau_h(x-he_i)}{2h}\\ \\ \qquad+f^{(0,1,0\dots0)}(x,z_0(x),\nabla z_0(x))\tau_h(x)\quad\text{for}\quad x\in\Omega_h.\end{cases}$$

For irregular points $x \in \Omega_{h,i}$ the changes corresponding to ([illegible]) have to be made. In (7.11) the $z_o(x) = \zeta_h(x)$ are known, the $\nabla z_o(x)$ may be computed by one of the methods defined above. In [6,8] we have shown that, under conditions satisfied here, we do not need in (7.11) $F'(z_o)$ exactly, but that we may pass over to an $F^*(z_o)$ which is obtained from $F'(z_o)$ by replacing $\nabla z_o(x)$ by

$$(7.13)\qquad \left(\frac{z_o(x+he_1) - z_o(x-he_1)}{2h}, \dots, \frac{z_o(x+he_n) - z_o(x-he_n)}{2h}\right) \approx \nabla z_o(x).$$

Now we are ready to formulate our discrete Newton method for the problem (1.1). Let $\varphi_h F$, as defined in (1.4)-(1.7), $F_1^*(z_o) := F'(z_o)$ and $F_2^*(z_o) := F^*(z_o)$, by replacing in $F'(z_o)$ $\nabla z_o(x)$ by the approximation in (7.13), $\varphi_h F_\nu^*(z_o)$ as in (7.11) and use one of the operators T in Method 1-4.

$$(7.14)\quad \begin{cases} \text{(i)} & \text{given the starting value } \zeta_{h,o} = \zeta_h \text{ such that} \\ & (\varphi_n F)\zeta_{h,o} = 0, \text{ define } z_{\ell-1} := T\zeta_{h,\ell-1} \text{ and} \\ & \text{compute } \zeta_{h,\ell}, \ell = 1,2,.. \text{ from} \\ \text{(ii)} & (\varphi_h F^*(z_o))(\zeta_{h,\ell} - \zeta_{h,\ell-1}) = - \begin{pmatrix} [illegible] \cap \\ 0 \text{ on } G_{h,n} \cap \partial\Omega \end{pmatrix} . \end{cases}$$

This method is called discrete Newton method. As indicated already in Pereyra-Proskurowski-Widlund [24] the order results for deferred corrections, als well as for discrete Newton methods and iterated defect corrections, are rather poor due to the fact that in (5.7) we have only $O(h^{k-1+\frac{1}{q}})$, instead of an $O(h^{k+1})$ which would correspond to the well known order results for $k = 0$ and $k = 1$.

Theorem 7.2: Let $k = 6$, $\lambda \geq 6$ in Method 1-4, and let the conditions in Theorem 5.3 be satisfied with $q = 3$ and $\alpha > 0.5$, so $z \in C^{8+\alpha}(\bar{\Omega})$. Further, let $\zeta_{h,1}$ be defined either as the first deferred correction or by (7.10) or (7.14). Then

$$(7.15)\qquad \zeta_{h,1} = \Delta_h z + O(h^{3.5}).$$

Further, if (6.1) is satisfied we have instead of (7.15)

$$(7.16)\qquad \zeta_{h,1} = \Delta_h\{z + h^4 e_{4,1}\} + O(h^{4.5})$$

If we compare the different types of defect and deferred corrections with the Richardson extrapolation, the main advantage of the extrapolation is to work up to order 5.5 compared to order 3.5 for the orther methods and to provide us with asymptotic upper and lower bounds for the exact solution (see Bulirsch-Stoer [13]). However, one has to pay for these advantages by much higher computational time than for discrete Newton methods or one of its equivalents. Expecially, for nonlinear problems (2.2) the discrete Newton methods yield the corrections in a relatively small amount of additional computations, by solving linear equations with a known matrix twice (see 8.). For those problems the discrete Newton methods are even superior to iterated defect and deferred corrections which furnish the corrections by solving nonlinear systems of the initial size with very good known starting values, whereas Richardson extrapolation needs the solution of nonlinear systems, with numbers of equations proportional to $1/h^n$.

8. Computational remarks:

In equation (2.6) the fraction $\alpha_{1,i}/\alpha_{o,i}$ will become very large if s_i is close to 1, that is, if the irregular point $x \in \Omega_{h,i}$ is close to the boundary. Therefore we will scale the equations in (2.6) always so, that the coefficient in the linear part of $\eta_h(x)$ becomes $2n\underline{a}$.

If (2.2) is linear we may use any good method to solve the linear system (2.4)-(2.6). The choice of the method very much depends on the properties of f, see e.g. Rosser [25] . In many cases it will be possible to get the corrections in the discrete Newton method (we choose $\nu = 2$ in (7.10) for such a case) by considerably less computational effort. That is, for example, true if the differential equation in (2.2) is the Poisson equation (treated in Pereyra-Proskurowski-Widlung [24] or the Helmholtz equation $\Delta z + cz = 0$, c = constant).For both equations one can very well use the fast Laplace solvers as described in [24] and then gets the corrections pretty cheaply.

If (2.2) is nonlinear, the system (2.4)-(2.8) will be nonlinear, too. In most cases one will then use the Newton method

$$(8.1)\qquad (\varphi_h F)'(\eta_{\ell-1})(\eta_\ell-\eta_{\ell-1}) = -(\varphi_h F)\eta_{\ell-1}, \quad \ell = 1,2,\ldots$$

to obtain the solution $\zeta_h = \lim\limits_{\ell\to\infty} \eta_\ell$ of (2.4)-(2.8). If an appropriate stoping criterium for the iteration (8.1) has been chosen, so $\zeta_h := \eta_\ell$, one may often use the already known matrix $(\varphi_h F)'(\eta_{\ell-1})$ for $\varphi_h(F^*(z_0))$ in (7.10). So, again, the corrections by the discrete Newton method are obtained relatively cheap.

Finally, we have pointed out already in 7. that for a proper choice of h the discrete Newton method may be executed on a fixed grid $\Gamma_{h,n}$ and (6.2) will usually be satisfied. So the chances are pretty good that a second correction improves the approximation. In that case it is worth-while to use $\lambda = k+1$ in Methods 1-4.

If (6.1) is satisfied, and the situation is smooth enough, it is advisable to take pretty big ($\lambda \geq 10$). Because of the loss of asymptotic terms in the methods discussed in Theorem 7.2 Richardson extrapolation might still be interesting for less smooth situations.

9. Numerical examples

As demonstrating examples we use boundary value problems, all on the same domain Ω, defined by the differential equations $F_i z = 0$, $i = 1,\dots,4$, and by the boundary values, which are the restrictions of the exact solutions z_i, $i = 1,\dots,4$, to the boundary $\partial\Omega$. So we have

$$(9.1)\begin{cases} \Omega: = \{(x,y)|(x-0.5)^2 + (y-1)^2 \leq (0.36)^2\} \;; \\ F_1 z: = z_{xx}+z_{yy} + 2z = 0, \; z(x,y) = \sin(xy); \\ F_2 z \;= z_{xx}+z_{yy} + 13z = 0, \; z(x,y)= \sin(2x+3y); \\ F_3 z \;= z_{xx}+z_{yy} + 2e^{-2z}=0, \; z(x,y)= \ln(x+y); \\ F_4 z \;= z_{xx}+z_{yy} -12z^2=0 \;, \; z(x,y) = (x+y)^{-2}. \end{cases}$$

With the discretization in §2 we obtain $\varphi_h F_1$ and $\varphi_h F_2$ as systems of linear, $\varphi_h F_3$ and $\varphi_h F_4$ as systems of nonlinear equations respectively. The first step in (7.10) and (7.14) is to solve $(\varphi_h F_i)\zeta_h = 0$. The nonlinear systems, for $i = 3,4$, are solved via the usual Newton method,see [8],(which is not to be mixed up with the discrete Newton method below). The matrices of all the problems above are positive definite and, for the special case of Ω in (9.1), symmetric. So we solve the linear equations by using SOR methods combined with techniques to estimate the optimal relaxation parameters. For the linear problems F_1 and F_2 these relaxation parameters have to be estimated only once, since the discretization $\varphi_h F_i$, i=1,2, and the discrete Newton method are based on the same matrix. For the nonlinear problems the matrix in the discrete Newton method is often replaced by the last matrix in Newton's method to solve the nonlinear system $(\varphi_h F_i)\zeta_h=0$, i=3,4. So the estimation for the optimal relaxation parameter in the last Newton step again may be used in the discrete Newton method. Therefore the iterations in the discrete Newton method are rather cheaply available.

In the following Table we give the errors $\| \zeta_{h,\nu}-\Delta_h z\|_2$ for $\nu = 0,1,2$ and for k = 3,...,6, where k is given in (2.5). The defects in the discrete

Newton methods are computed with methods 3 and 4, see §7. In method 3 we use overlapping polynomials with $\kappa = 2$ and in method 4 we only document those results, where we have used polynomials of degree 9 in (7.7) to compute provisional values for the outer points which we need for the symmetric divided differences. Because of (7.15) we have

$$\zeta_{h,\nu-1} - \Delta_h z \approx \zeta_{h,\nu-1} - \zeta_{h,\nu}$$

Therefore the correction $\zeta_{h,\nu-1}-\zeta_{h,\nu}$ may be used as an estimation for the error $\zeta_{h,\nu-1}-\Delta_h z$, especially $\zeta_{h,2}-\zeta_{h,3}$ has been used to estimate $\zeta_{h,2}-\Delta_h z$. The numbers q are defined as quotients between these estimated errors and the real errors.

The following numbers are computed on a UNIVAC 1108 with double precision (about 18 digits). I want to thank cand. math. H. Offermann, who did the computations.

		Example 1				Example 2			
		method 3		method 4		method 3		method 4	
k	ν	error	q	error	q	error	q	error	q
3	0	1.3,-06	1.0	1.3,-06	1.0	8.2,-05	1.0	8.2,-05	1.0
	1	1.6,-09	1.1	2.3,-08	1.0	3.0,-07	1.0	4.3,-07	1.0
	2	1.7,-10	0.6	7.6,-10	0.8	1.3,-08	0.6	4.2,-08	0.9
4	1	4.2,-10	1.0	2.0,-08	1.0	1.4,-07	1.0	4.8,-09	1.0
	2	8.6,-12	0.9	2.4,-10	1.0	2.3,-09	0.8	3.4,-09	1.3
5	1	4.3,-10	1.0	2.0,-08	1.0	1.4,-07	1.0	6.3,-08	1.0
	2	1.2,-12	1.0	2.4,-10	1.0	4.6,-10	1.0	6.4,-10	2.0
6	1	4.2,-10	1.0	2.0,-08	1.0	1.5,-07	1.0	6.5,-08	1.0
	2	9.6,-13	1.0	2.4,-10	1.0	2.8,-10	1.0	3.8,-10	2.4
		Example 3				Example 4			
3	0	2.5,-06	1.0	2.5,-06	1.0	2.4,-05	1.0	2.4,-05	1.0
	1	6.6,-09	1.1	2.3,-08	0.9	7.7,-08	1.0	3.2,-07	1.0
	2	1.1,-09	0.6	2.5,-09	0.8	1.6,-08	0.6	3.5,-08	0.7
4	1	2.8,-09	1.0	1.8,-09	1.0	6.4,-08	1.0	3.6,-08	1.1
	2	1.2,-10	0.8	1.6,-10	0.8	2.4,-09	0.7	3.0,-09	0.6
5	1	3.2,-09	1.0	3.5,-10	1.0	7.6,-08	1.0	4.2,-09	0.9
	2	1.3,-11	0.8	2.0,-11	0.4	2.6,-10	0.7	1.1,-09	0.6
6	1	3.3,-09	1.0	1.9,-10	0.9	7.8,-08	1.0	3.4,-09	1.1
	2	5.2,-12	1.0	3.4,-11	0.7	3.2,-10	1.1	1.5,-09	1.1

Table

From this table we see that, even for k = 3, the first correction is worthwhile and for larger values of k, especially k = 6, two corrections considerably improve the approximation and that, expecially in method 3, the third correction still gives an excellent estimation for the error.

REFERENCES

[1] Agmon, S., A. Douglis, L. Nirenberg: Estimates near the boundary for solutions of elliptic partial differential equations satisfying general boundary conditions I, Comm. Pure Appl. Math. 12, 623-727 (1959).

[2] Ames, W. F.: Nonlinear partial differential equations in engineering, Academic Press, New York-London 1965 .

[3] Bers, L.: On mildly nonlinear partial differential equations of elliptic type, J. Res. Nat. Bur. Standards, 51, 229-236 (1953).

[4] Böhmer, K.: Spline Funktionen, Teubner Verlag, Stuttgart, 1974.

[5] Böhmer, K.: A defect correction method for functional equations, in: Approximation theory, Bonn 1976, Eds. R. Schaback and K. Scherer, Springer Lecture Notes in Mathematics, #556, 16-29 (1976).

[6] Böhmer, K.: Fehlerasymptotik von Diskretisierungsverfahren und ihre numerische Anwendung, Univ. Karlsruhe, Institut für Praktische Mathematik, Interner Bericht Nr. 77/2 (1977).

[7] Böhmer, K.: Defekt-Korrektur-Methoden für das zentrale Euler-Verfahren bei Randwertproblemen gewöhnlicher Differentialgleichungen , in ISNM, Birkhäuser-Verlag, Basel, Eds.: Albrecht, Collatz, Hämmerlin, 1978.

[8] Böhmer, K.: Discrete Newton methods and iterated defect corrections, I General theory, submitted to Numer. Math.,II. Initial and boundary value problems in ordinary differential equations, to appear.

[9] Böhmer,K.: Defekt-Korrektur-Methoden für Integralgleichungen zweiter Art, to appear in ZAMM.,Tagungsband (1978).

[10] Böhmer, K.: Asymptotic expansions for the discretization error in Poisson's equation on general domains, in "Multivariate approximation theory", Eds.W. Schempp, K. Zeller, ISNM 51,30-45 (1979).

[11] Bramble, J.H., B.E. Hubbard: Approximation of derivatives by finite difference methods in elliptic boundary value problems,Contributions to Differential Equations,3,393-410 (1964).

[12] Brandt, A.: Estimates of difference quotients of solutions of Poisson type differential equations, Math. Comp. 20, 473-499 (1966).

[13] Bulirsch, R., J. Stoer: Fehlerabschätzungen und Extrapolation mit rationalen Funktionen bei Verfahren vom Richardson-Typus, Numer. Math. 6, 413-427 (1964).

[14] Collatz, L. Bemerkungen zur Fehlerabschätzung für das Differenzenverfahren bei partiellen Differentialgleichungen, Z. Angew. Math. Mech., 13, 56-57 (1933)

[15] Frank, R.: Schätzungen des globalen Diskretisierungsfehlers bei Runge-Kutta-Methoden, ISNM 27, 45-70 (1975).

[16] Frank, R.: The method of iterated defect-correction and its application to two-point boundary value problems, Part I /II, Numer. Math. 25, 409-419 (1976) /[illegible], 407-420 (1977).

[17] Frank, R., J. Hertling, C.W. Ueberhuber : Iterated defect correction based on estimates of the local discretization error, Report Nr. 18/76 des Instituts für Numerische Mathematik, Technische Universität Wien (1976).

[18] Frank, R., J. Hertling, C. W. Ueberhuber: A new approach to iterated defect corrections, to appear.

[19] Frank, R., J. Hertling,:Die Anwendung der iterierten Defektkorrektur auf das Dirichletproblem, Beiträge zur Numer. Math. 8,...(1978)

[20] Gerschgorin, S.: Fehlerabschätzung für das Differenzenverfahren zur Lösung partieller Differentialgleichungen, Z. Angew. Math. Mech. 10, 373-382 (1930).

[21] Gerschgorin, S.: Über die Abgrenzung der Eigenwerte einer Matrix, Izv. Akad. Nauk. SSSR Ser. Mat. 7, 749-754 (1931).

[22] Pereyra, V.: Iterated deferred corrections for nonlinear operator equations, Numer. Math., 10, 316-323 (1967).

[23] Pereyra, V.: Highly accurate numerical solution of casilinear elliptic boundary value problems in n dimensions, Math. Comp., 24, 771-783 (1970).

[24] Pereyra, V., W. Proskurowski and O. Widlund: High order fast Laplace solvers for the Dirichlet problem on qeneral regions, Math. Comp., 31, 1-16 (1977).

[25] Rosser, J. B.: The direct solution of difference analogs of Poisson's equation, MRC Technical Summary Report #797, University of Wisconsin-Madison (1967).

[26] Stetter, H. J.: Analysis of discretization methods for ordinary differential equations, Springer-Verlag, Berlin-Heidelberg-New York, 1973.

[27] Stetter, H. J.: Economical global error estimation, in "Stiff Differential Systems", R. A. Willoughby, ed., 245-258, Plenum Press, New York - London, 1974.

[28] Wasow, W.: Discrete approximations to elliptic differential equations, Z. Angew. Math. Phys. 6, 81-97 (1955).

[29] Zadunaisky, P. E.: A method for the estimation of errors propagated in the numerical solution of a system of ordinary differential equations, in "The Theory of orbits in the solar system and in stellar system ", Proc. of Intern. Astronomical Union, Symp. 25, Thessaloniki, Ed. G. Contopoulos, 1964.

[30] Zadunaisky, P. E.: On the estimation of errors propagated in the numerical integration of ordinary differential equations, Numer. Math. 27, 21-40 (1976).

KB:db

(14) MRC-TSR-2042

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 2042	2. GOVT ACCESSION NO. AD-A083 824	3. ~~RECIPIENT'S CATALOG NUMBER~~ (9) Technical
4. ~~TITLE (and Subtitle)~~ (6) HIGH ORDER DIFFERENCE METHODS FOR QUASILINEAR ELLIPTIC BOUNDARY VALUE PROBLEMS ON GENERAL REGIONS.		5. ~~TYPE OF REPORT & PERIOD COVERED~~ Summary Rep~~ort~~, ~~no specific reporting period~~
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) (10) Klaus Böhmer		8. CONTRACT OR GRANT NUMBER(s) (15) DAAG29-75-C-0024 ~~BO 62271~~
9. PERFORMING ORGANIZATION NAME AND ADDRESS Mathematics Research Center, University of Wisconsin 610 Walnut Street Madison, Wisconsin 53706		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 7 - Numeri~~cal~~ Analysis
11. CONTROLLING OFFICE NAME AND ADDRESS See Item 18 below		12. ~~REPORT DATE~~ (11) Feb~~ruary~~ 1980
		13. NUMBER OF PAGES ~~43~~ (12) 47
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

U. S. Army Research Office
P.O. Box 12211
Research Triangle Park
North Carolina 27709

Deutsche Forschungsgemeinschaft
Bonn, Germany

University of Karlsruhe
Karlsruhe, Germany

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

elliptic differential equations, difference equations, asymptotic error expansion, defect corrections, stability, heat transfer

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Stability and convergence for a difference method for quasilinear elliptic boundary value problems are proved. Asymptotic expansions of the discretization error, basic for Richardson extrapolation, are established. The general theory of "discrete Newton methods" and "iterated defect corrections via neighboring problems" ~~[6,8]~~ and Pereyra's deferred corrections ~~[22]~~ are used to derive different high order methods. Some special cases and computational problems are pointed out and numerical tests are included.

221200